2019

Comparing Gut Microbiome and Virome in the Breast Milk- and Formula-fed Late Preterm Infants

Zyi Wang
South Dakota State University

COMPARING GUT MICROBIOME AND VIROME IN THE BREAST MILK- AND FORMULA-FED LATE PRETERM INFANTS

BY

ZIYI WANG

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Biological Sciences

Specialization in Microbiology

South Dakota State University

2019

COMPARING GUT MICROBIOME AND VIROME IN THE BREAST MILK- AND FORMULA-FED LATE PRETERM INFANTS

Ziyi Wang

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Biological Sciences and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Shin-Yi Marzano, Ph.D.
Thesis Advisor — Date 7/25/19

Xiuqing Wang, Ph.D.
Thesis co-Advisor — Date 7/25/19

Volker Brozel, Ph.D.
Head, Department of Biology and Microbiology — Date 07/26/2019

Dean, Graduate School — Date 7-26-19

ACKNOWLEDGEMENTS

Firstly, I would like to show my deepest gratitude to my thesis advisor Dr. Shin-Yi Lee Marzano and my supervisor Dr. Xiuqing Wang for their support and valuable guidance throughout my master's in science research. They taught me not only about research technology and skills, but also the ability to explore the unknown. I couldn't finish this thesis without them. I would also like to thank my other committee members, Drs. Joy Scaria and Hesham Fahmy for their valuable input to my project.

My appreciation goes to Drs. Jessica White and Richard Vo at Sanford Health, who supported me on the clinical aspect to have a more complete understanding of my research. I would also like to thank the nurses who have helped me with the sample collection.

Last but not least, I would like to thank all my colleagues in the Marzano laboratory, especially Achal Neupane who have helped me a lot on data analysis. Additionally, I shall extend my thanks to Dr. Runan Yao who have helped me on the use of R program. And finally, I would like to acknowledge the support provided by my roommate Yijie Fan and my parents during the preparation of my thesis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

COMPARING GUT MICROBIOME AND VIROME IN THE BREAST MILK- AND FORMULA-FED LATE PRETERM INFANTS

ZIYI WANG

2019

The neonatal intestinal microbiome consists of all microorganisms in the gut. Although the microbiome is critical to human health and disease, its colonization remains incompletely understood, particularly in the preterm infant. We aimed to characterize the intestinal bacteria microbiome and virome in moderate-to-late preterm infants. We hypothesized that the bacteria microbiome and virome differs between breast milk and formula-fed infants. We collected stool samples from twenty infants born between 32 0/7 and 36 6/7 weeks gestation. Samples were collected after infants reached full volume enteral feedings. Ten infants were breast milk fed and ten received infant formula. DNA and RNA were extracted from fecal samples and sequenced using amplicon sequencing of 16S rRNA seq and RNAseq. 16S results showed that breast milk and formula-fed infants had similar bacterial diversity. *Firmicutes* were found in all samples and constituted the predominate phylum in most of infants regardless of nutrition. Breast milk-fed infants have 18% and 15% of *Veillonella* and *Escherichia*, respectively compared to 10% and 0.02% in formula-fed. 18% and 17% of *Streptococcus* and *Klebsiella* in formula group, but 9% and 8% in breast milk group. Overall, the abundant of *Propionibacterium* was significantly higher in breast milk-fed infants than formula fed. These results were basically consistent with metatranscriptome results except breast milk-fed group had more *Streptococcus than* formula group. For virome composition, we

identified three different bacteriophages and discovered that the read counts of *Siphoviridae* were significantly higher in formula fed infants from metatranscriptome results (p-value = 0.002). Based on sample analysis from these twenty preterm infants, we concluded that the preterm intestinal microbiome is altered by diet. While microbial diversity was similar between breast milk- and formula-fed infants, the predominant bacteria differed. The abundance of *Siphoviridae* seems to be related to formula-fed. Our results provide new knowledge on diet affection of moderate to late preterm infants in both bacteria microbiome and virome aspect, and two different methods, 16S sequencing and RNAseq, commonly used to study intestinal microbes were compared to provide a reference for selection.

CHAPTER ONE: LITERATURE REVIEW

1. INTRODUCTION

Human gut microbiome composition is very complex. The density of microbial populations in human colon may reach up to 10^{11}cells/g content, and may constitute 1-2 kg of weight[2]. Most of the microorganisms in the gut are considered to be beneficial to the human body. Some of them even can participate in the digestion process. Many previous studies found that although there are huge individual differences between different healthy adults, they do share some core microbiome which are considered to play an key role in maintaining human health[3-5]. In healthy adults' intestines, *Firmicutes* and *Bacteroidetes* are the most abundant phylum[4, 6, 7]. On the contrary, it's known that some microbiome is also involved in some diseases such as non-alcoholic fatty liver disease (NAFLD)[8, 9], obesity[10] and cardiovascular disease[11, 12].

The composition of infant's gut microbiome is highly dynamic and changes as infants grow[13]. This is contrary to the gut microbiome of adults, which are usually stable. There are still lots unanswered questions regarding how the early microbiome colonization of the newborn intestine is formed, and whether healthy infants share a core microbiome community like adults. Previous studies found that infant gut microbiome can be influenced by many factors, and this will lead to great individual differences in gut microbiome in their first year[2, 14]. Besides that, there are a lot of evidence supports that a relevance between infant gut microbiome and many diseases exists, including necrotizing enterocolitis (NEC), irritable bowel syndrome, inflammatory bowel disease, etc.[15-17]. Therefore, it is critical to understand the mode of first colonization of microbiome in infants and how different factors may affect this.

2. COMPOSITION OF INFANT GUT MICROBIOME

2.1. Bacterial microbiome

A few decades ago, it was thought that the gut of newborn infants is sterile, but many studies have overturned this hypothesis[17, 18]. Jiménez et al. isolated some bacteria from first meconium samples of healthy infants[19]. In regard to how a fetus gets its first gut bacteria, many studies suggest that fetuses obtain their first gut bacteria from their mother's amniotic fluid[19-22]. The amniotic fluid is considered as the first fluid into fetus gut by swallowing during gestational period[22]. The specific mechanism of bacterial transfer from mother-to-fetus is not quite clear still, but a labeled *Enterococcus fecium*, which was orally inoculated in pregnant mice, was found in their baby mice meconium[20]. Despite this, rapid establishment and development of gut microbiome colonization in infant occurs postpartum. The composition of infant gut bacterial microbiomes was found to differ very widely and affected by many factors, including gestational age, nutrition, antibiotic treatment, method of delivery, genetic, gender and environments expose[16, 17].

2.1.1. Gestational age effect

Gestational age is one of the key factors to affecting the first batch of colonizers in infant gut. Full-term infants gut often have abundant *Bifidobacterium* and *Lactobacillus* which are considered as beneficial bacteria[23, 24]. Preterm infants, however, were found to have a significant delay of *Bifidobacterium* and *Lactobacillus* colonization in their early days of life[25, 26]. Instead, they are more likely colonized dominantly by potentially pathogenic bacteria such as *Staphylococcus*, *Escherichia coli*, *Enterobacteriales*, *Bacteroides*, *Clostridiales* or *Klebsiella*[27-32]. Preterm infants always have lower

diversity and slower growing of bacterial than full-term babies[27]. Many evidence showed that richness and diversity of bacterial microbiome in infants increased with age from birth to 2 or 3 years old and it becomes close to adults [13, 33, 34].

2.1.2. Nutrition effect

Diet, as an important factor affecting infancy, has attracted the attention of many researchers. Breast milk and formula milk are among the few choices of nutrition for early infancy. Formula milk has an optimal nutritional ratio and has been proven to help promote a higher growth rates of preterm infants[35, 36]. However, there is growing evidences to show that breast milk contains immunoglobulins, anti-inflammatory factors, lactoferrin and hormones[37-41] which may be directly or indirectly related with promoting neonatal health, such as decreasing the risk of necrotizing enterocolitis (NEC)[35, 42-44], decreasing the rate of obesity in childhood[45], promoting mental development in preterm infants[46] etc. Gut microbiome of breast milk-fed infants are predominantly *Bifidobacterium* with minor components of *Lactobacillus, Streptococci* and *Staphylococci*[30, 47-50], which corresponds to a healthy composition of human breast milk[49, 51-53]. Formula-fed infants are more likely be co-dominated by *Bifidobacterium* and *Bacteroides,* minor components are *Escherichia coli*, *Clostridia, Enterococci*[30, 47, 48, 50].

2.1.3. Antibiotic use

The effect of antibiotics on early intestinal colonization in infants is enormous, because antibiotics will affect the richness and diversity of the microbiome community. Neonatal sepsis, which is a common disease with high mortality rate in newborns especially in preterm infants, generally was caused by bacteria infection like *Group B*

Streptococcus (GBS), *Escherichia coli* or *Staphylococcus*[54, 55]. Ampicillin with gentamicin is the common prescribed antibiotics to prevent the disease[55]. Fouhy et al. found that infants who were taken a ampicillin with gentamicin treatment had a greater proportion of *Proteobacteria* but a lesser abundance of *Actinobacteria* and *Lactobacillus*[56]. Another study found more *Enterobacteriaceae*, *Enterococcus* and *Staphylococcus* in the gut of infants who received the ampicillin and gentamicin[57]. In addition, the richness of *Clostridium* and *Lactobacillus* was limited by penicillin or penicillin plus gentamicin antibiotic treatment in infant gut, but it had little influence on *E. coli*[27]. *Klebsiella* and *E. coli* show potential resistance to amoxicillin plus netilmicin treatment in human gut. Similarly, *Staphylococci* shows the potential resistance to amoxicillin plus netilmicin and cefotaxime[58].

2.1.4. Delivery methods

Delivery method is also a confirmed factor influencing the microbial colonization of the infant gut. Cesarean delivery (CsD) infants were more likely have abundance of *Staphylococcus*, *Corynebacterium* or *Propionibacterium spp.*, which are more similar to bacterial composition of their mother's skin. Vaginally delivered (VD) babies' gut were detected to have *Lactobacillus*, *Prevotella* and *Sneathia spp.,* which are more similar to their mother's vaginal environment[59]. Besides that, VD infants are more likely predominated by *Bifidobacterium* which is lacks in CsD infants[60], but CsD infants seems are more easily to carry *Enterobacteria and Clostridium*[61, 62]. However, it seems that the delivery method has little effect on long-term intestinal colonization. Hill et al. found that although full-term CsD infants have more Firmicutes and less Actinobacteria

than full-term VD infants at first week, the intestinal composition of these two groups of infants tends to have no significant difference from eight week to one year[63].

2.1.5. Gender effect

According to the study by Cong *et al.*[47], gut microbiome community of males are more likely to be dominated by *Enterobacteriales*, and those of females are more likely *Clostridiales* and *Lactobacillales*. Besides that, males have lower α-diversity (0.34±0.3) than females (0.49±0.23) within first 10 days after birth. Then, the α-diversity of men and women tends to be not significantly different.

2.2. Virome

2.2.1. Bacteriophage in the infant gut virome

In the earliest life of infants, very few amount of viruses are found in their gastrointestinal (GI) tract compared with bacteria[13, 16]. A study showed that the most abundant viruses in infant gut virome were bacteriophages, specifically phages under *Caudovirales* order[13]. However, the virome progressed to be predominated by *Microviridae* family over two years[13, 64]. On the other hand, *Microviridae* were also detected as the dominant bacteriophage in the adult gut virome[65]. In addition, the richness and diversity of bacteriophage in the infant gut virome are higher than in adults, and will decrease with age[13, 16].

2.2.2. Eukaryotic viruses in infant gut virome

In addition to bacteriophages, the presence of both DNA eukaryotic virus and RNA eukaryotic viruses in infant gut is extremely low[16, 64]. The detected families are *Anelloviruses*, *Caliciviruses*, *Astroviruses*, *Picornaviruses*, *Adenoviruse*s, *Proviruses*,

Enteroviruses, *Parechoviruses*, *Tombamoviruses* and *Sapoviruses*[13, 16, 66]. Among them, *Anelloviruses* has been detected multiple times by different studies with a higher richness[13, 66, 67].

2.2.3. Factors influencing the infant gut virome

Unlike bacteria, the relationship between viruses in infants and diet is somewhat ambiguous. One study reported that infant fecal samples shared 30% of viral contigs with their mothers' breast milk[68]. In contrast, Breitbarta *et al.* reported that the most abundant viruses in infant fecal samples cannot be found in breast milk or formula food[64]. Another study focused on the relationship between adult human gut virome and their diet and obtained the similar results. People who ate the same food, didn't have the same viruses in their fecal samples[69]. This could indicate that we don't obtain our stable GI tract viruses directly from our diet. However, in the same study, one group of people who changed their food choice had changes in their gut virome. This indicated that diet may have some influence on the human gut virome.

Genetics are known to play a role. There are some studies that focused on twins. Identical twin infants were confirmed to have similar compositions of gut virome than fraternal twins, then nonrelated infants[13, 16, 70]. This could be due to the genetics of infants. Interestingly, there are some significant differences in gut virome between co-twin adults[71]. Not only genetics is responsible for gut virome composition, but environmental factors certainly play a role at the same time. This leads to highly dynamic and huge differences in human gut virome within an individual over time, and among individuals due to the complex and uncontrollable environment.

3. CONCLUSION

Infant gut microbiome is extremely dynamic and is affected by multiple factors at the same time[13, 16, 64]. Factors include but are not limited to gestational age, delivery method, nutrition, gender, gene and environment[13, 16-18]. The richness and diversity of bacteria will increase with age, but bacteriophage will decrease. So, compare with adult gut microbiome, infants have lower bacterial diversity and higher bacteriophage diversity[13, 16]. However, the gut microbiome of infant has lower stability than adults. It means that infant gut microbiome exhibits more individual differences.

CHAPTER TWO:

GUT BACTERIA IN THE BREAST MILK- AND FORMULA-FED PRETERM INFANTS

1. INTRODUCTION

1.1. Overview of gut microbiome

The intestinal microbiome has been gradually attracting scientific interest because it has been found to be closely related to human health and diseases. Although individual differences exist, some bacteria, such as *Firmicutes* and *Bacteroidetes*, have been found to dominate the intestines of healthy adults and are considered the bacterial core to maintain human health[3, 4, 6, 7]. Unlike adults, colonization of the infant gut flora is dynamic and unstable. There are still many unanswered questions regarding how the early microbiome colonization of the newborn intestine forms and whether healthy infants share a core microbiome community like adults. Previous studies found that the infant gut microbiome can be influenced by many factors that will lead to great individual differences in the gut microbiome in their first year [2, 14]. It is known that many factors, including genetic factors [70] and non-genetic/environmental factors such as nutrition, antibiotic use, and delivery method could affect the infants' microbiome composition and richness[2, 14, 16].

Some evidence supports a possible association between the infant gut microbiome and disease, such as necrotizing enterocolitis (NEC), inflammatory bowel disease (IBD), irritable bowel syndrome, etc. [15-17]. Many previous studies focused on the microbiome in the first month or the first year of full-term infants[13, 16, 17, 24, 48, 61, 68, 72]. Some studies have been performed to examine the intestinal microbial diversity and richness in very preterm infants[30, 32, 73-75]. Preterm infants are babies who born before 37 weeks of

gestation. The three sub-categories of preterm infants based on gestational age are: extremely preterm (less than 28 weeks), very preterm (28-32 weeks) and late preterm (32-37 weeks). To our knowledge, few studies focused on the gut microbiome of late preterm infants. Previous studies suggest that *Bifidobacteria* has higher abundance in the gut of full-term infants than preterm infants[23, 29, 50].

1.2. Nutrition effect on preterm infants

Nutrition is undoubtedly a key factor affecting the composition of infant intestinal microbes. During infancy, especially in preterm infants, the only source of nutrition is breast milk and/or formula. Formula milk is the preferred alternative when mothers cannot produce enough breast milk. Formula is considered a balanced nutrient source and could reduce care burdens on the mothers. A few studies claim that formula could promote higher growth rates of preterm infants than breast milk [35, 36]. In recent years, however, breastfeeding has been valued by more and more people because of the protective maternal antibodies and complex community of bacteria provided by breast milk[37-41]. Some studies pointed out the potential role of breast milk in promoting neonatal health, such as reducing the rate of obesity in childhood, [45] and promoting mental development in preterm infants [46]. In addition, breast milk is also believed to be responsible for decreasing the risk of NEC in preterm infants [35, 42, 43]. Gewolb et al. showed that the alpha-diversity of the gut microbiome began to show significant differences between breast milk-fed preterm and formula-fed infants when they reach 30 days old, but not when 10 and 20 days old [76]. Cong et al. sequenced stool samples from preterm infants during their first 30 days and showed that the gut microbiome in breast milk-fed preterm infants was more diverse than that of formula-fed preterm infants [47].

Other studies found that *Bifidobacteria* more easily colonizes the gut of breast milk-fed full-term newborns than the gut of formula-fed infants, possibly because it is already present in breast milk [24, 48]. However, most of the studies couldn't find an enrichment of *Bifidobacteria* in preterm neonates' intestines regardless of nutrition source [27, 47, 76, 77]. Instead, *Enterobacteriaceae* and *Klebsiella*, from the *Proteobacteria* phylum, [28, 31, 76, 78] and *Enterococcus*, *Staphylococcus*, *Streptococcus*, *Clostridium*, *Lactobacillales* and *Veillonella*, from the *Firmicutes* phylum, [28, 47, 76-78] are the main abundant bacteria in preterm infants. Thus, the purpose of this study was to characterize the difference of intestinal bacterial microbiomes between breast milk- and formula-fed moderate-to-late preterm infants.

1.3. Objective and hypotheses

The objective of the study was to characterize the intestinal bacterial microbiome in moderate-to-late preterm infants. We hypothesized that the microbiome differs between breast milk- and formula-fed infants.

2. MATERIAL AND METHODS

2.1. Sample collection.

All the experiments that involved human subjects in this study were approved by the Sanford Health IRB, IRB ID: STUDY00000829IRA. Twenty healthy preterm infants born between 32 0/7 to 36 6/7 weeks gestation were recruited to the study. Fecal specimens were collected after infants reached full volume enteral feedings in Sanford Children's MB2 Clinic, Sioux Falls, from March 30th, 2017 to December 10th, 2017. All infants were assigned a unique ID number, and the bioinformation of infants is shown in Table 1. Ten of the twenty infants were breast milk-fed and ten were formula-fed.

Although mothers were very active in breastfeeding, breast milk-fed infants still received formula support. However, formula-fed infants did not receive any breast milk supplement. All mothers took antibiotics. Eleven of twenty infants received Ampicillin/Gentamicin antibiotics to combat suspected symptoms of sepsis. Six of twenty were born through vaginal delivery and fourteen were born via Cesarean section (C-section). The race of six infants were described as Native American and the rest of infants were Caucasian. Additionally, there are five pairs of twins: infants #7 and #8; #9 and #10; #18 and #19; #21 and #22; #23 and #24. For each infant, we chose two time points for sample collection. The first time points were collected at an average of 15 days after birth and the average of the second time points was 16 days after birth. Each collected sample was stored in DNA/RNA shield tubes (Zymo Research), which contain DNA/RNA shield that could reduce nucleic acid degradation. All samples were stored at 4^{o}C.

2.2. DNA/RNA extraction

Total DNA and RNA were simultaneously extracted from the fecal samples that stored in DNA/RNA shield tubes by using DNA/RNA Mini Kit (ZRC188678, ZymoBIOMICSTM) with a small modification. Briefly, we removed 800 µl of shield from the lysis tube provided by the kit and added 1ml of fecal samples into the tube. The remaining steps were performed as described by the manufacturer's instructions. The concentrations of the extracted DNA/RNA were quantified with both Qubit 3.0 Fluorometer (Catalog Number Q33216, Invitrogen™) and NanoDrop™ Spectrophotometer (Catalog Number ND-2000, Thermo Scientific™).

Table 1. Demographic and clinical information of infants enrolled in the study.

Infants number (#)	Gender	Nutrition	Antibiotic using	Delivery type	Twin pair
1	F	Formula	0	Cs	
2	M	Breast milk	1	Cs	
3	F	Breast milk	0	Cs	
7	F	Formula	0	V	A
8	M	Formula	0	V	A
9	M	Breast milk	0	Cs	B
10	M	Breast milk	0	Cs	B
11	F	Breast milk	1	V	
13	F	Breast milk	0	Cs	
15	M	Formula	0	V	
16	F	Breast milk	1	V	
18	M	Formula	1	Cs	C
19	F	Formula	1	Cs	C
20	F	Breast milk	1	V	
21	F	Breast milk	1	Cs	D
22	F	Breast milk	1	Cs	D
23	M	Formula	1	Cs	E
24	M	Formula	1	Cs	E
25	F	Formula	1	Cs	
27	M	Formula	0	Cs	

Gender: F = Female, M = Male
Antibiotics using: 0 = Non-antibiotics, 1 = Receive Ampicillin/Gentamicin antibiotic treatment
Delivery type: Cs = Cesarean section, V = Vaginal delivery
Twins: A, B, C, D, E five pairs of twins were included.

2.3. 16S rRNA sequencing

The 16S rRNA sequencing of V3-V4 segments was done using a MiSeq Next Generation Sequencer by the University of Minnesota Genomics Center[79].

2.4. Library preparation and RNA sequencing

Genomic RNA of sixteen infants (#2, #3, #9, #10, #11, #13, #20, #21, #22 from the breast milk-fed group and #1, #7, #8, #18, #19, #23, #24 from the formula-fed group) with two separated time points were extracted. Among them, RNA of two time points of

infants #1, #2, #3, #7, #8, #9, #10, #11 and #13 were used to prepare the RNAseq libraries separately, but, for infants #18, #19, #20, #21, #22, #23 and #24, we pooled the RNA of both two time points together to prepare for the RNAseq libraries. A total of 12 µl RNA was used to deplete the human rRNAs using a rRNA Depletion Kit (Human/Mouse/Rat) (E6310L, NEBNext®) by following the manufacturer's instructions. Library preparation was performed by using a Directional RNA Library Prep Kit for Illumina kit (E7760S, NEBNext®).

2.5. Metatranscriptomic analysis

The Metatranscriptome data was trimmed by Trimmomatic[80] first to cut the adaptors off and then DIAMOND[81] was used to aligned DNA reads and MEGAN6[82] to analysis. The bubble chart, bar charts, radial trees, rarefaction plot and eggNOG gene functional analysis were done by the MEGAN6 program.

2.6. Statistical analysis.

Unequal variance (Welch) unpaired t test using GraphPad Prism software was performed to examine the statistical significance of gestational age, birth weight, the first and second sample collection times between the two groups of enrolled infants.

For amplicon sequencing of the bacterial microbiome, an average of 44,558 paired-end reads per sample were obtained. Sequences were filtered at Phred 33 and demultiplexed by Quantitative Insights Into Microbial Ecology 2 (QIIME2)[83]. Then we discarded all sequences that had a length of over 300 nucleotides. The identity threshold of 97% operational taxonomic units (OTU) table was made using the Greengenes database. Jaccord of beta diversity analysis and ANCOM were performed through QIIME2. Additionally, another method of 16S analysis, USEARCH [84], has been adopted

and used in this analysis. This is a novel sequence analysis tool that has gradually been accepted by more researchers. Raw data was trimmed and filtered with expected errs < 1 by USEARCH. The 97% of OTU table, alpha and beta diversity, taxonomy with rdp_16s_v16.fa database, frequency distance of OTUs between two groups and random forest of machine learning to get informative OTUs were performed by using the UPARSE pipeline of USEARCH. Boxplots were done by using R program.

3. RESULTS

3.1. *Demographic information of enrolled infants*

An unequal variance (Welch) unpaired t test was done between ten breast milk-fed and ten formula fed preterm infants; it showed no significant differences in gestational age, birth weight, first sample and second sample age between the two groups of enrolled infants (Table 2). Therefore, no potential confounding factors between the treatment groups existed.

3.2. *Effect of nutrition on the diversity and richness of gut bacterial population*

3.2.1. Alpha-diversity and Beta diversity of 16S sequences

Based on the OTU table obtained from the UPARSE pipeline of the USEARCH program, alpha diversity and beta diversity between the two treatment groups were compared. Alpha diversity, the average species diversity in a sample, could be an important indicator to distinguish the differences between the groups. Overall, there were no significant differences in bacterial alpha diversity between breast milk- and formula-fed infants (Figure 1A, Supplementary Table 3). Nevertheless, the medium value of diversity of breast milk-fed infants is slightly higher than formula-fed infants, based on the Shannon index, and there was a larger variation in the formula-fed group.

In addition to alpha diversity, beta diversity, which is the ratio between regional and local species diversity, is another important indicator to determine whether different nutrition types could affect the infants' gut microbiome. The two separate clusters of different nutrition can be observed from the 3D jaccard plot of Figure 1B, which indicates that the beta diversity of bacteria composition in breast milk and formula fed infants are dissimilar enough to cluster separately.

Table 2. **Population demographics.** Other conditions of two diet group infants were listed and a t test of each condition was done between breast milk and formula fed groups. All the p values show that there are no significant differences in each potential factor between two diet groups.

	Breast milk	**Formula**	**p value**
Gestational age(weeks)	33.9 ± 1.7	34.1 ± 2	0.556
Birth weight(grams)	2050 ± 640	2380 ± 670	0.509
First sample age(days)	16.7 ± 9.7	15.4 ± 14.0	0.986
Second sample age(days)	20.0 ± 17.0	22.5 ± 17.5	0.907
Antibiotic exposure	60%	50%	
Vaginal delivery	30%	30%	
Gender(male)	30%	60%	
Twins	40%	60%	
Proton Pump Inhibitor	0%	0%	

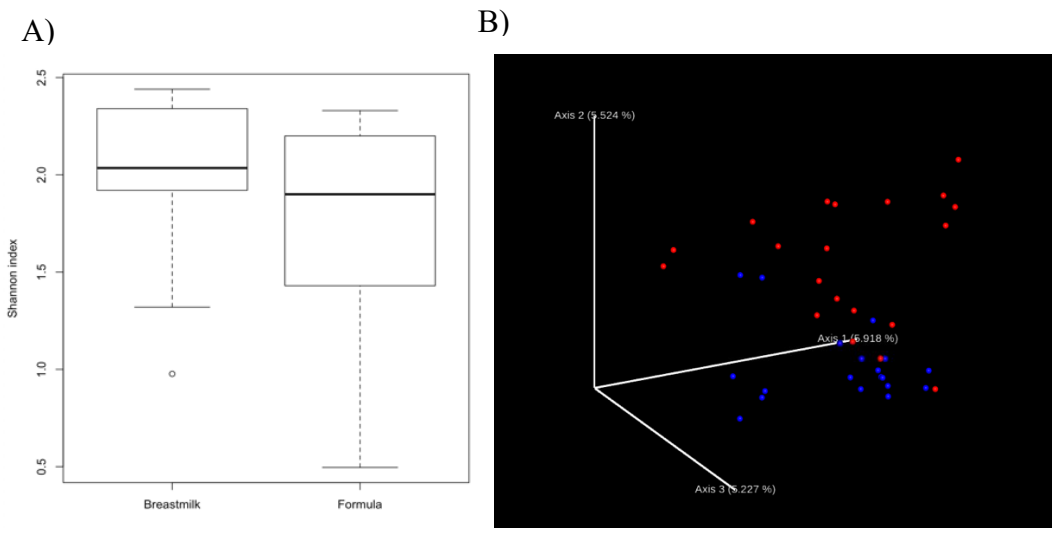


Figure 1. Alpha and beta diversity of different nutrition. (A) Shannon index of alpha diversity shows no significant difference between breast milk and formula-fed infants. P-value > 0.05. (B) Jaccord plot of beta diversity presented by EMPeror between two nutrition types that clustered separately. Red: Breast milk group. Blue: Formula group.

3.2.2. Taxonomic analysis of 16S data

3.2.2.1. Bacteria composition of different nutrition types at the phylum level

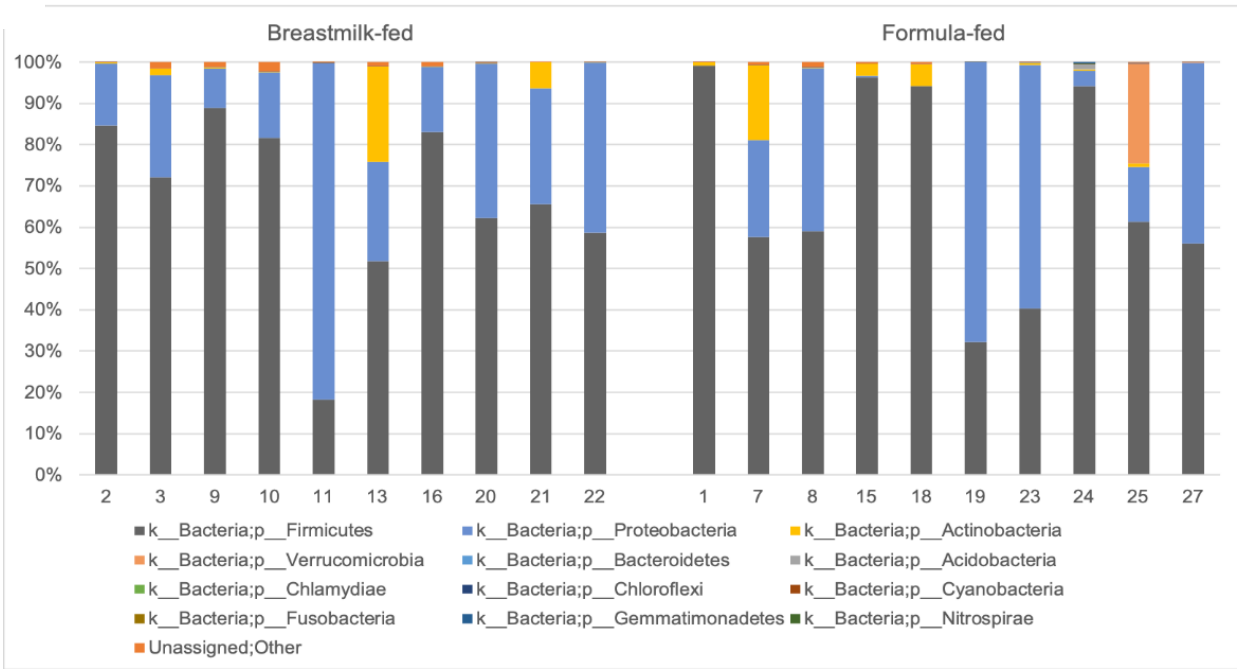
As the compositional difference was found in the beta diversity, OTUs at the phylum level were first profiled to reveal that *Firmicutes* was the most abundant bacteria phylum in the majority of breast milk-fed and formula-fed preterm infants with an average of 67% and 69% of the composition, respectively (Fig 2A). *Proteobacteria* was the second most abundant bacteria with a composition of 29% and 25% for the breast milk and formula group, respectively. Interestingly, *Proteobacteria* was the most dominant bacteria in infants, #11 in the breast milk group and #19 and #23 in the formula group, suggesting other factors at play. Also, *Actinobacteria* was observed in only some infants. There were no significant differences in the top three dominant bacteria phyla between breast milk- and formula-fed preterm infants (Figure 2B).

3.2.2.2. Bacteria composition of different nutrition types at the genus level

Analysis of the OTUs at the genus level using QIIME2 (Fig 3A, 3B) did not reveal a single dominant genus present in either breast milk-fed infants or formula-fed infants. The most abundant bacteria in breast milk-fed infants were the *Enterobacteriaceae* family and *Veillonella* genus at 16.6% and 15.9%, respectively. The formula group, however, showed more diversity and variation in dominant bacteria genus between the two sampling times (Fig 3A). For example, *Lactobacillus* was dominant in baby 15 at the first sampling time point but changed to *Veillonella* later; whereas *Klebsiella* was dominant in baby 19 at both sampling time points. Further, baby 23 was dominated by *Staphyloccus* at the second sampling time point but *Streptococcus* was dominant in baby 24 initially and was replaced by *Enterococcus* during the second sampling time point. Finally, baby 27 was dominated by *Haemophilus* initially and then by *Streptococcus*. The overall bacterial compositions were relatively stable between the two sampling time points of each infant. However, we observed differences in dominant genus in some infants in such a short sampling interval, suggesting the dynamic characteristics of gut microbiome in preterm infants, specifically in the formula-fed group. From the boxplot, we observed that there are significantly more *Enterobacteriaceae* family in the breast milk group but a less *Enterococcus* than the formula group (Figure 3B). Sintax taxonomy pie charts from the USEARCH pipeline showed the bacterial genera present in two groups was consistent with the results from QIIME2 (Figure 3C, 3D). The *Veillonella* genus and *Escherichia/Shigella* genus, which belongs to *Enterobacteriaceae* family, have the most richness in breast milk-fed preterm infants, at 18.4% and 15.2% respectively. They were followed by *Staphylococcus* (10.6%), *Clostridium* (9.6%), *Enterococcus*

(9.6%) and *Streptococcus* (9.3%). In formula-fed preterm infants, *Streptococcus* (18.6%) and *Klebsiella* (17.4%) were the most abundant, followed by *Enterococcus* (12.4%), *Staphylococcus* (10.7%) and *Veillonella* (10.4%). Comparing Fig 3C and Fig 3D, breast milk-fed infants have much more richness of *Escherichia* and *Veillonella* than the formula-fed group, but much less *Klebsiella* and *Streptococcus.* In addition, *Akkermansia* was uniquely present in the formula group, and so was *Lactococcus* in the breast milk group. It is worth noting that *Propionibacterium* shows 1.2% in breast milk-fed preterm infants at genus level, which is revealed by QIIME2 ANCOM analysis, and that it has significantly greater abundance in the breast milk group than the formula group (Fig 4).

A)



-

B)

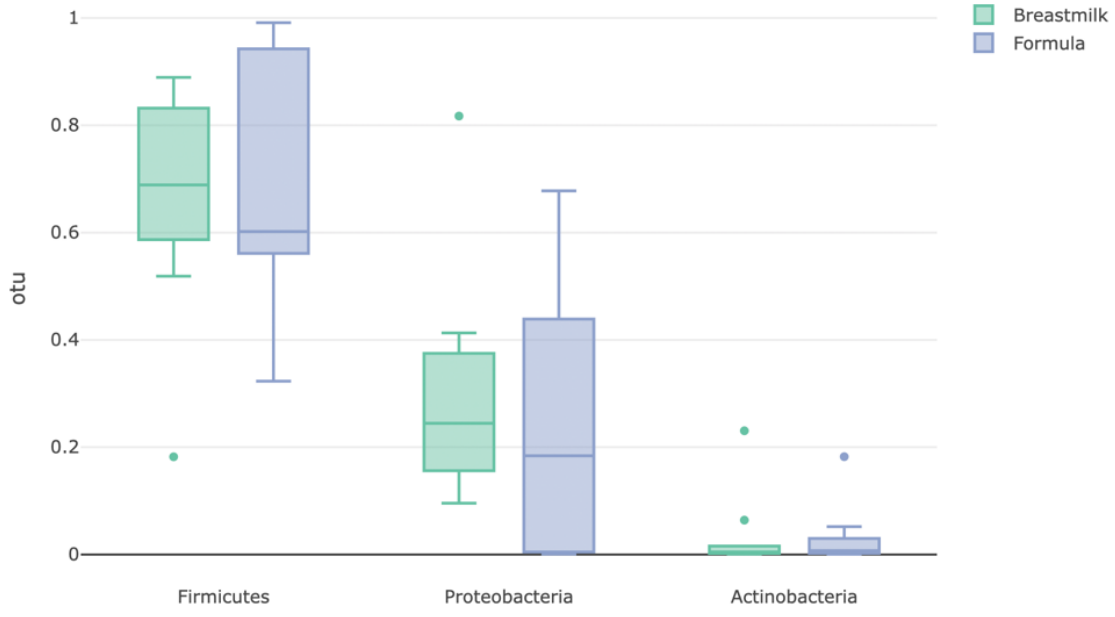
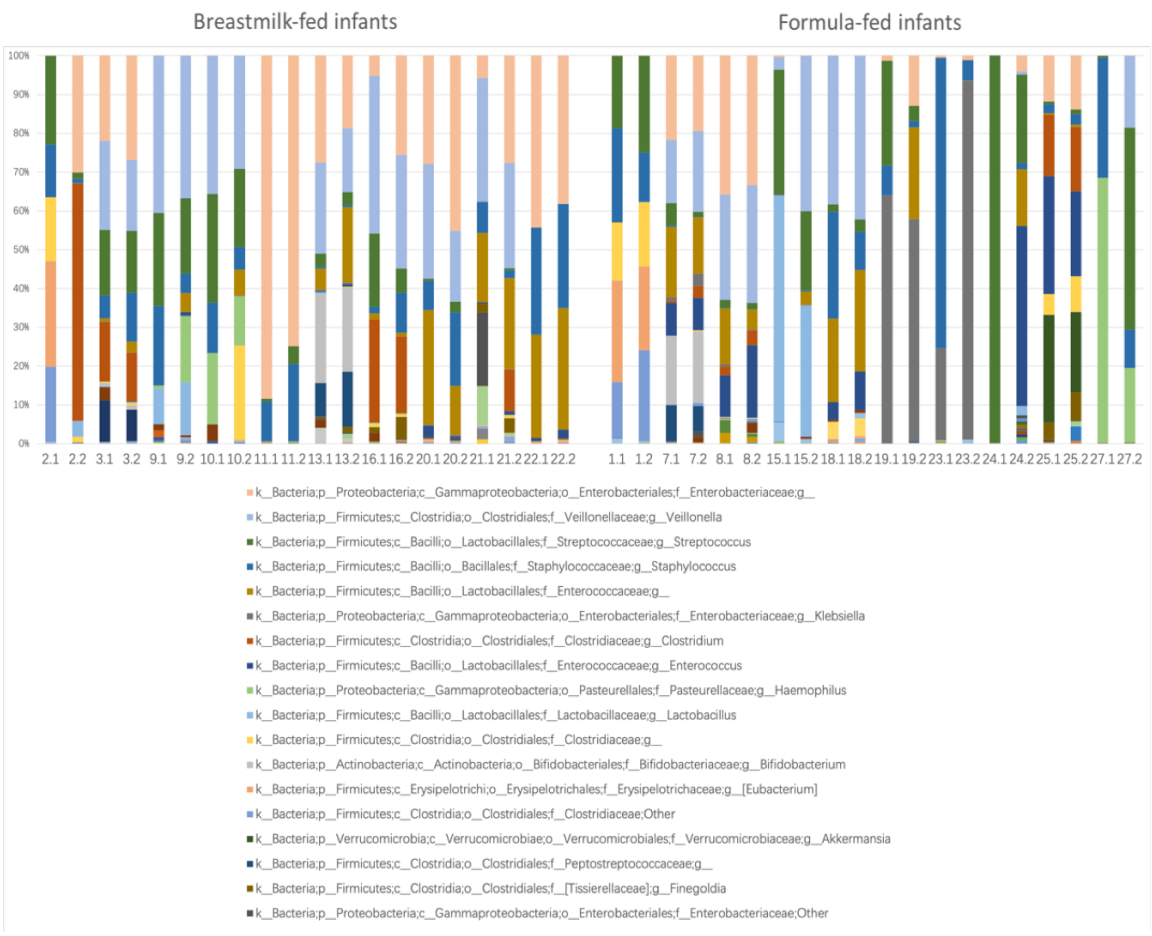
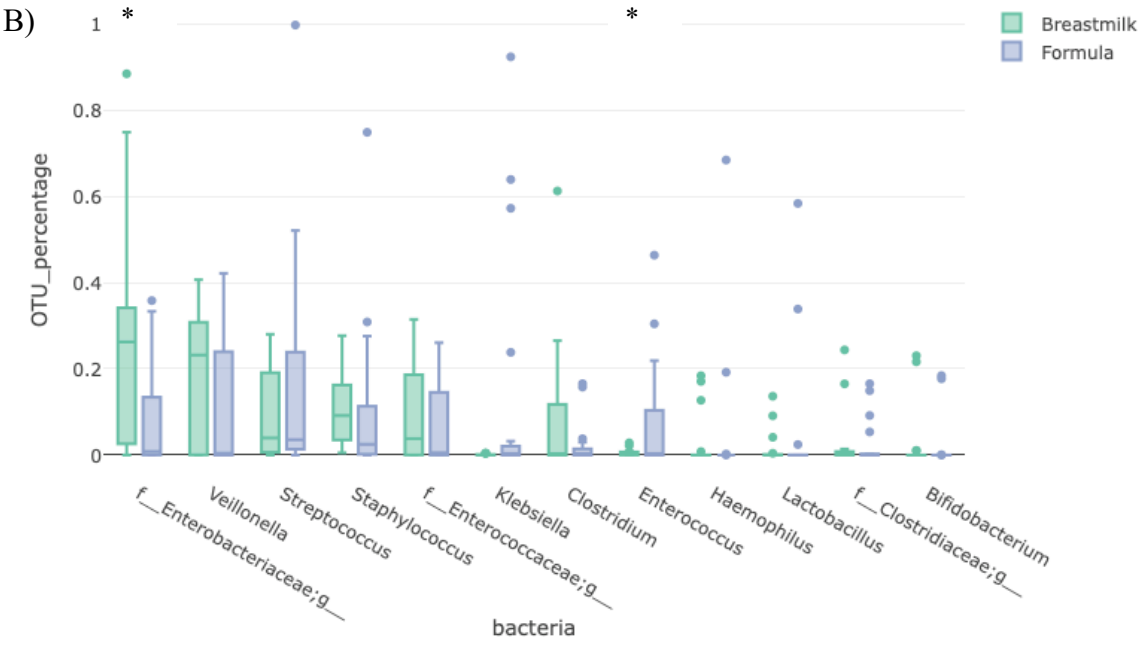


Figure 2. Bacteria composition of breast milk-fed and formula-fed preterm infants in phylum level. (A) Bar chart of bacteria composition and proportion of breast milk and formula fed preterm infants in phylum level. Two different time points were merged together. (B) Boxplot of the top three most abundant bacteria in breast milk and formula fed preterm infants in phylum level.

A)


Breastmilk-fed infants
Formula-fed infants
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
2.1 2.2 3.1 3.2 9.1 9.2 10.1 10.2 11.1 11.2 13.1 13.2 16.1 16.2 20.1 20.2 21.1 21.2 22.1 22.2
1.1 1.2 7.1 7.2 8.1 8.2 15.1 15.2 18.1 18.2 19.1 19.2 23.1 23.2 24.1 24.2 25.1 25.2 27.1 27.2
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Klebsiella
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__[Eubacterium]
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;Other
k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Finegoldia
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;Other

B)


*
*
Breastmilk
Formula
1
0.8
0.6
0.4
0.2
0
OTU_percentage
f__Enterobacteriaceae;g__
Veillonella
Streptococcus
Staphylococcus
f__Enterococcaceae;g__
Klebsiella
Clostridium
Enterococcus
Haemophilus
Lactobacillus
f__Clostridiaceae;g__
Bifidobacterium
bacteria

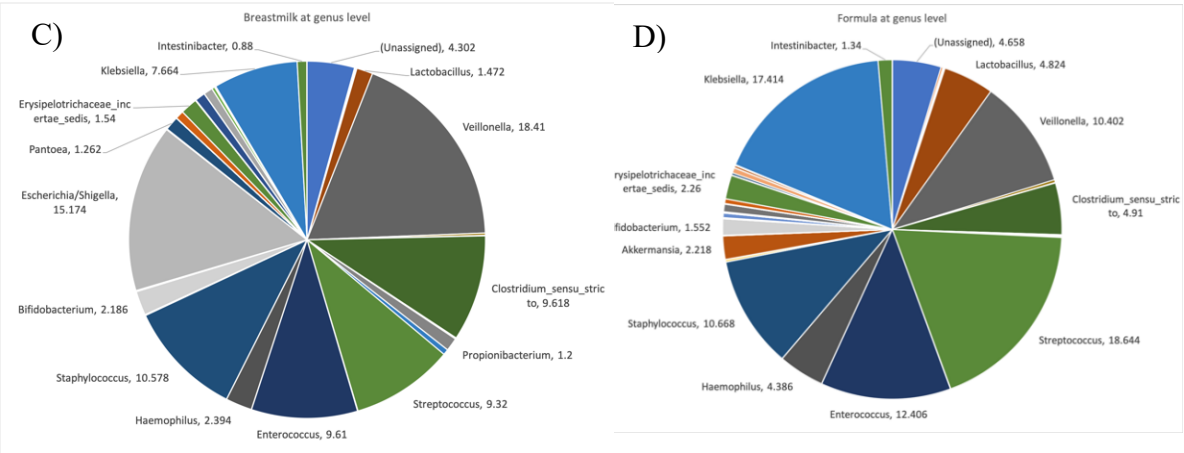


Figure 3. Bacteria composition of Breast milk-fed and Formula-fed preterm infants in genus level. (A) Bacteria composition and proportion of each breast milk and formula fed preterm infant with two time points in genus level. (B) Boxplot of top 12 bacterial genus of different nutrition from QIIME2. * There were significant differences of family *Enterobacteriaceae* and genus *Enterococcus* between the two nutrition groups, p-value < 0.05. (C) Taxonomy pie charts of the proportion in breast milk-fed group in genus level from UPARSE. The number of each sector is the percentage of that bacteria. (D) Taxonomy pie charts of the proportion in formula-fed group in genus level from UPARSE. The number by each section refers to the percentage of that bacteria.

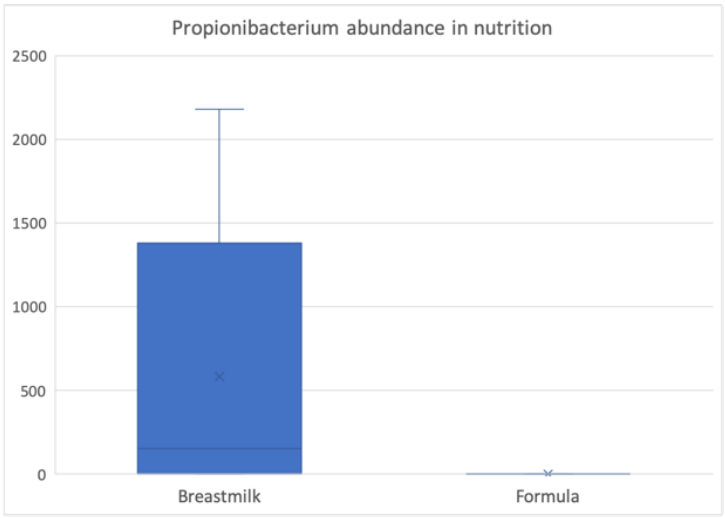


Figure 4. Boxplot showing the abundance of *Propionibacterium* in breast milk-fed and formula-fed infants. The abundance of *Propionibacterium* is significantly higher in breast milk-fed than formula-fed infants.

3.2.3. Informative OTUs identified to differentiate different nutrition types

Random forest classifier in the UPARSE pipeline agreed with the ANCOM analysis that OTU19-*Propionibacterium* was informative that the counts can be used effectively to separate breast milk and formula groups. Cross-checking Table 3 with the Table 4, which also a part of results from the UPARSE pipeline, OTU 34 and OTU 77, which both represent *Nitrososphaera*, were found at the 21st and 13th frequency distance of OTUs, respectively. OTU 4, *Staphylococcus*, was the 9th frequency distance of OTUs and more abundant in breast milk-fed infants. OTU 26, *Rothia*, was the 10th rank of frequency and higher in the formula-fed group. Therefore, not only *Propionibacterium* mentioned above, but also *Nitrososphaera*, *Staphylococcus* and *Rothia*, could use to distinguish between nutrition sources.

Table 3. Top 10 informative OTUs identified by random forest classifier[1] that can be used to distinguish breast milk-fed from formula-fed infants. The OTUs were listed by orders.

OTU number	Importance *	Frequency category**	Taxonomy predictions
Otu19	0.0137	Breast milk	d:Bacteria,p:Actinobacteria,c:Actinobacteria,o:Actinomycetales,f:Propionibacteriaceae, g:Propionibacterium
Otu34	0.0095	Formula	d:Archaea,p:Thaumarchaeota,o:Nitrososphaerales,f:Nitrososphaeraceae,g:Nitrososphaera
Otu282	0.00933	Formula	d:Others
Otu4	0.00887	Breast milk	d:Bacteria,p:Firmicutes,c:Bacilli,o:Bacillales,f:Staphylococcaceae,g:Staphylococcus
Otu26	0.00785	Formula	d:Bacteria,p:Actinobacteria,c:Actinobacteria,o:Actinomycetales,f:Micrococcaceae,g:Rothia
Otu77	0.00636	Formula	d:Archaea,p:Thaumarchaeota,o:Nitrososphaerales,f:Nitrososphaeraceae,g:Nitrososphaera
Otu329	0.00573	Formula	d:Bacteria,p:Actinobacteria,c:Actinobacteria
Otu69	0.00549	Formula	d:Bacteria,p:Acidobacteria,c:Acidobacteria_Gp4
Otu3	0.00517	Breast milk	d:Bacteria,p:Proteobacteria,c:Gammaproteobacteria,o:Enterobacteriales,f:Enterobacteriaceae, g:Escherichia/Shigella
Otu37	0.00507	Breast milk	d:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Streptococcaceae,g:Streptococcus

* The "Importance" value was calculated by random forest classifier through UPARSE pipeline to represents the importance of each OTU. The higher importance value an OTU has, the more informative it has to distinguish between two different groups.
** Frequency category indicates the group showing the frequency of a particular OTU.

Table 4. The top 30 frequency distance rank of OTUs in nutrition.

Rank	MinGini	MaxAUC	AUC*	ScoreA	ScoreG	Med	Breast milk**	Name
1	0.2308	0.85	0.835	4	4	0	+	**Otu19**
2	0.32	0.8	0.76	0.5	0.5	1	-	**Otu130**
3	0.3333	0.75	0.75	0.5	0.5	0	-	**Otu282**
4	0.3333	0.75	0.675	2.30E+02	2.30E+02	0	+	**Otu3**
5	0.3333	0.75	0.75	1	1	0	+	**Otu63**
6	0.3333	0.75	0.75	0.5	0.5	0	-	**Otu571**
7	0.3333	0.75	0.75	0.5	0.5	0	-	**Otu49**
8	0.3333	0.75	0.775	1.5	1.5	0	-	**Otu117**
9	0.3626	0.75	0.61	2.50E+02	2.50E+02	4.60E+02	+	**Otu4**
10	0.3626	0.75	0.74	1.5	1.5	0	-	**Otu26**
11	0.3737	0.75	0.75	0.5	0.5	0	-	**Otu150**
12	0.375	0.7	0.7	4.5	4.5	0	-	**Otu689**
13	0.375	0.7	0.72	1.5	1.5	0	-	**Otu77**
14	0.375	0.7	0.7	0.5	0.5	0	-	**Otu366**
15	0.375	0.7	0.7	0.5	0.5	0	-	**Otu329**
16	0.375	0.7	0.7	0.5	0.5	0	-	**Otu167**
17	0.375	0.7	0.7	0.5	0.5	0	-	**Otu248**
18	0.375	0.7	0.7	0.5	0.5	0	+	**Otu1010**
19	0.4048	0.7	0.675	0.5	0.5	0	-	**Otu250**
20	0.4048	0.7	0.65	1.5	1.5	0	-	**Otu90**
21	0.4048	0.7	0.695	6.5	5.5	5	-	**Otu34**
22	0.4048	0.7	0.675	0.5	0.5	0	+	**Otu14**
23	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu825**
24	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu238**
25	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu281**
26	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu359**
27	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu392**
28	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu259**
29	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu10**
30	0.4118	0.65	0.65	0.5	0.5	0	-	**Otu23**

* AUC = Area under the ROC curve

** (+) or (-) indicates whether high or low abundance implies the positive category.

3.3. *Metatranscriptome results*

Comparing the QIIME2 and UPARSE results of 16S sequencing, metatranscriptomic analysis at the genus level of sixteen infants showed a roughly consistent composition of microbiome, but different richnesses (Figure 5A). *Streptococcus* and *Clostridium* are the most abundant bacteria genus in the breast milk group (Figure 5B), as well as *Escherichia*, *Veillonella* and *Bifidobacterium* (Figure 5D). In the formula group, the most abundant bacteria genus are *Enterococcus* and *Streptococcus*, followed by *Bifidobacterium*, *Clostridium* and *Pantoea* (Figure 5C, 5E). Although both the breast milk-fed and formula-fed groups have a high abundance of

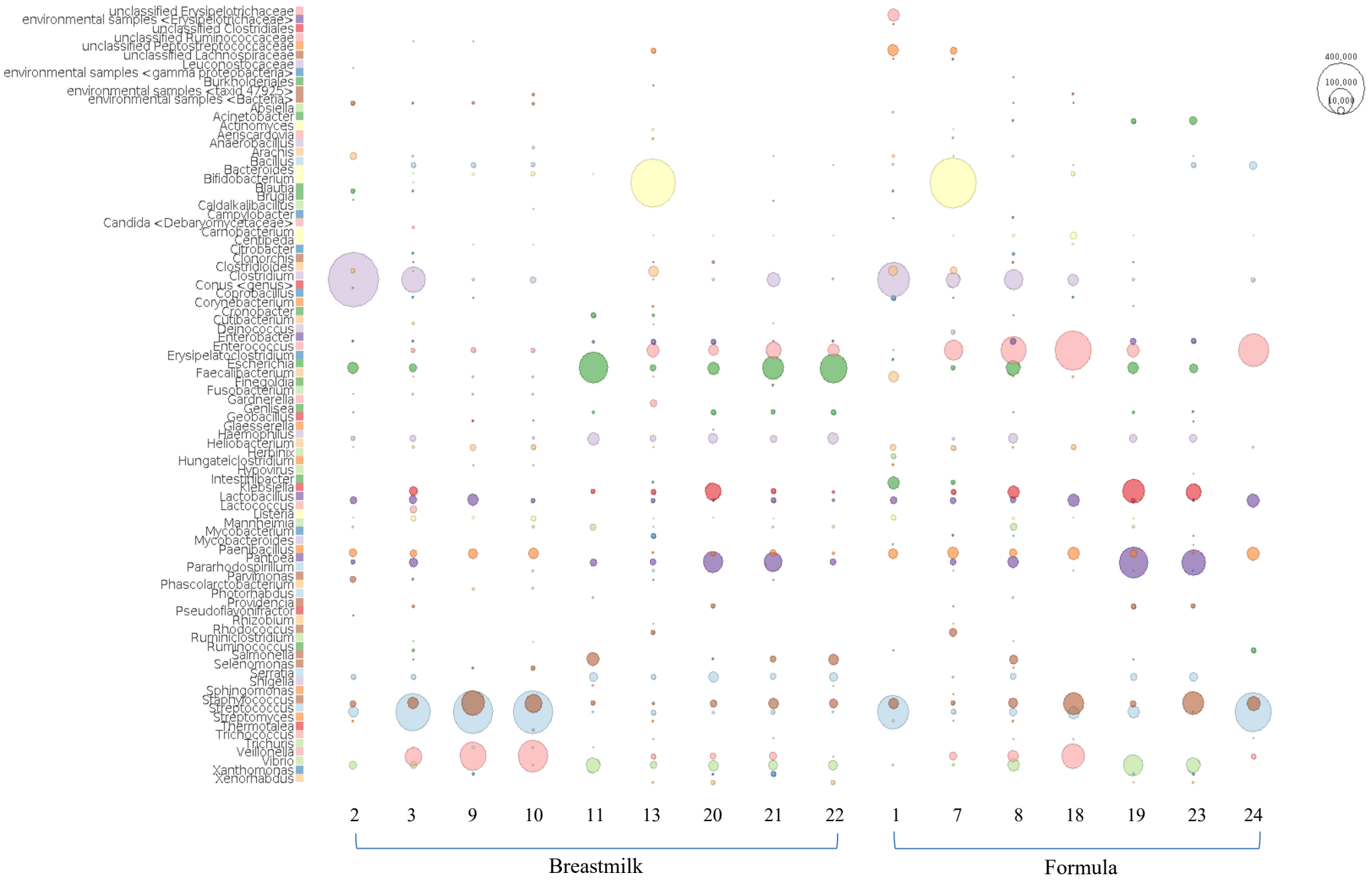
Bifidobacterium, we can observe that *Bifidobacterium* was only dominant in #13 and #7, respectively, in both groups (Figure 5B, 5C). Overall, in addition to *Clostridium*, *Streptococcus*, *Veillonella, Bifidobacterium*, *Enterococcus*, *Escherichia* and *Staphylococcus,* that showed abundance in both 16S sequencing and metatranscriptomics analyses, *Vibrio* and *Paenibacillus* were specifically detected to be abundant by metatranscriptome analysis (Figure 5D, 5E).

The taxonomy rarefaction plot shows that the alpha diversity of all sixteen infants was similar (Figure 6). The PCoA plot of eight infants with two time points separated shows that beta-diversity was similar within two time points for each infant but dispersed between infants (Figure 7). Among them, #7 and #8 are a pair of twins and #9 and #10 are another pair. We can also observe that except 10-2, which could be considered an outlier, the beta-diversity within twins was more similar than unrelated infants.

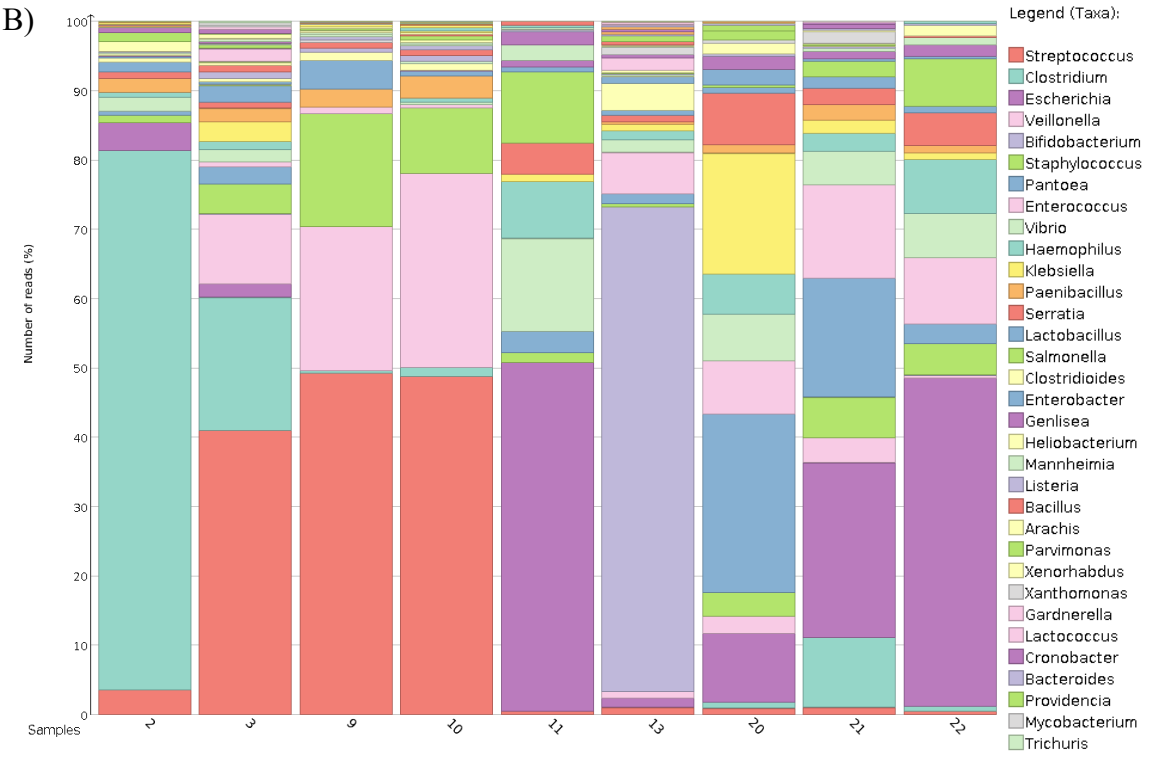
The metatranscriptome data was subsequently analyzed through gene functional analysis in MEGAN6. The first nine columns of InterPro2GO result represent breast milk-fed infants (#2, #3, #9, #10, #11, #13, #20, #21, #22), followed by seven columns of formula-fed infants (#1, #7, #8, #18, #19, #23, #24) (Figure 8A). Comparing the breast milk group and formula group in Figure 11A, the formula group appears to have more hits of the three main function groups detected, especially in cellular component function. Based on the InterPro2GO bar chart of breast milk-fed infants (Figure 8B) and formula-fed infants (Figure 8C), the most abundant functions in both two groups are metabolic process and catalytic activity. But, other than that, there were several noticeable differences between the two groups. The breast milk group has more ion binding and nucleic acid binding functions detected than the formula group and less intrinsic

component of membrane and transport functions. We also tried the eggNOG and SEED databases (Supplementary figure 1 and 2) to analyze gene function. There, however, no very obvious difference was observed.

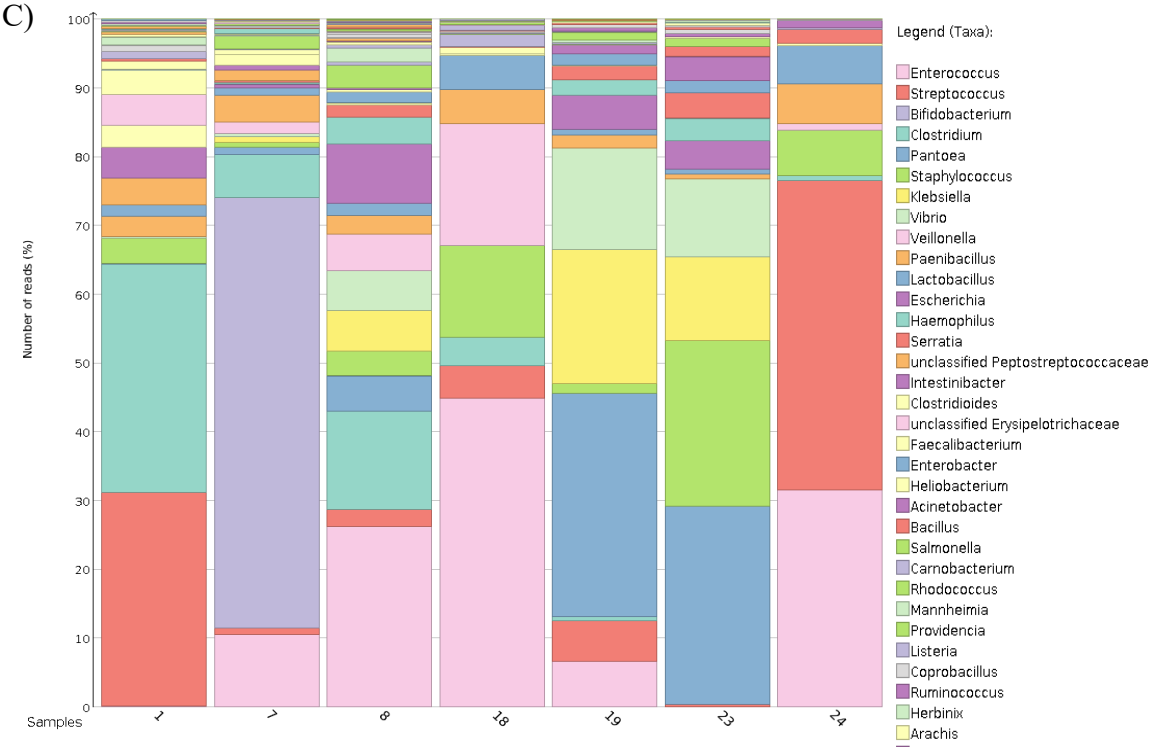
A)


unclassified Erysipelotrichaceae
environmental samples <Erysipelotrichaceae>
unclassified Clostridiales
unclassified Ruminococcaceae
unclassified Peptostreptococcaceae
unclassified Lachnospiraceae
Leuconostocaceae
environmental samples <gamma proteobacteria>
Burkholderiales
environmental samples <taxid 47925>
environmental samples <Bacteria>
Absiella
Acinetobacter
Actinomyces
Aeriscardovia
Anaerobacillus
Arachnis
Bacillus
Bacteroides
Bifidobacterium
Blautia
Brucia
Caldalkalibacillus
Campylobacter
Candida <Debaryomycetaceae>
Carnobacterium
Centipeda
Citrobacter
Clonorchis
Clostridioides
Clostridium
Conus <genus>
Coprobacillus
Corynebacterium
Cronobacter
Cutibacterium
Deinococcus
Enterobacter
Enterococcus
Erysipelatoclostridium
Escherichia
Faecalibacterium
Finegoldia
Fusobacterium
Gardnerella
Genlisea
Geobacillus
Glaesserella
Haemophilus
Heliobacterium
Herpinix
Hungateiclostridium
Hypovirus
Intestinibacter
Klebsiella
Lactobacillus
Lactococcus
Listeria
Mannheimia
Mycobacterium
Mycobacteroides
Paenibacillus
Pantoea
Pararhodospirillum
Parvimonas
Phascolarctobacterium
Photorhabdus
Providencia
Pseudoflavonifractor
Rhizobium
Rhodococcus
Ruminiclostridium
Ruminococcus
Salmonella
Selenomonas
Serratia
Shigella
Sphingomonas
Staphylococcus
Streptococcus
Streptomyces
Thermotalea
Trichococcus
Tuchuris
Veillonella
Vibrio
Xanthomonas
Xenorhabdus
400,000
100,000
10,000
2
3
9
10
11
13
20
21
22
1
7
8
18
19
23
24
Breastmilk
Formula

B)


Number of reads (%)
Samples
2
3
9
10
11
13
20
21
22
Legend (Taxa):
Streptococcus
Clostridium
Escherichia
Veillonella
Bifidobacterium
Staphylococcus
Pantoea
Enterococcus
Vibrio
Haemophilus
Klebsiella
Paenibacillus
Serratia
Lactobacillus
Salmonella
Clostridioides
Enterobacter
Genlisea
Heliobacterium
Mannheimia
Listeria
Bacillus
Arachis
Parvimonas
Xenorhabdus
Xanthomonas
Gardnerella
Lactococcus
Cronobacter
Bacteroides
Providencia
Mycobacterium
Trichuris

C)


Number of reads (%)
Samples
1
7
8
18
19
23
24
Legend (Taxa):
Enterococcus
Streptococcus
Bifidobacterium
Clostridium
Pantoea
Staphylococcus
Klebsiella
Vibrio
Veillonella
Paenibacillus
Lactobacillus
Escherichia
Haemophilus
Serratia
unclassified Peptostreptococcaceae
Intestinibacter
Clostridioides
unclassified Erysipelotrichaceae
Faecalibacterium
Enterobacter
Heliobacterium
Acinetobacter
Bacillus
Salmonella
Carnobacterium
Rhodococcus
Mannheimia
Providencia
Listeria
Coprobacillus
Ruminococcus
Herbinix
Arachis

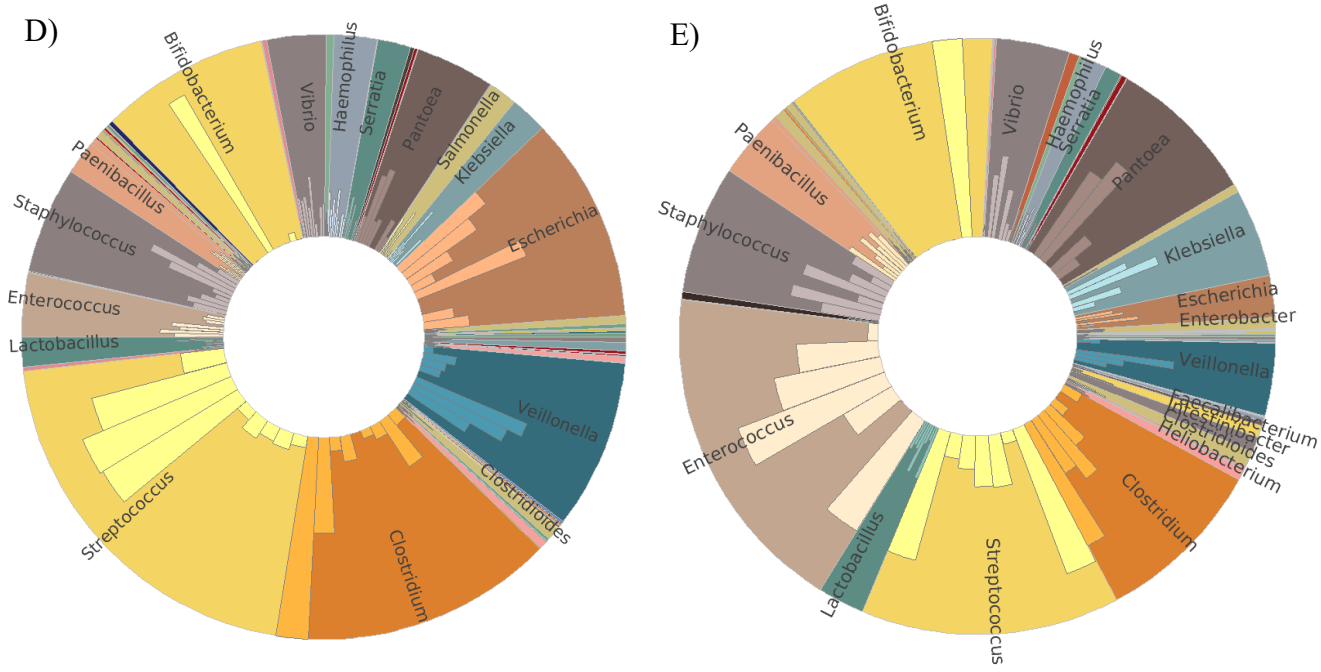


Figure 5. Charts of sixteen infants metatranscriptomic results at genus level. (A) Bubble chart of sixteen infants metatranscriptomic results. (B) Bar chart of metatranscriptomic results in breast milk-fed infants (#2, #3, #9, #10, #11, #13, #20, #21, #22). (C) Bar chart of metatranscriptomic results in formula-fed infants (#1, #7, #8, #18, #19, #23, #24). (D) Radial tree of nine breast milk-fed infants (#2, #3, #9, #10, #11, #13, #20, #21, #22) metatranscriptomic result. (E) Radial tree of seven formula-fed infants (#1, #7, #8, #18, #19, #23, #24).metatranscriptomic result.

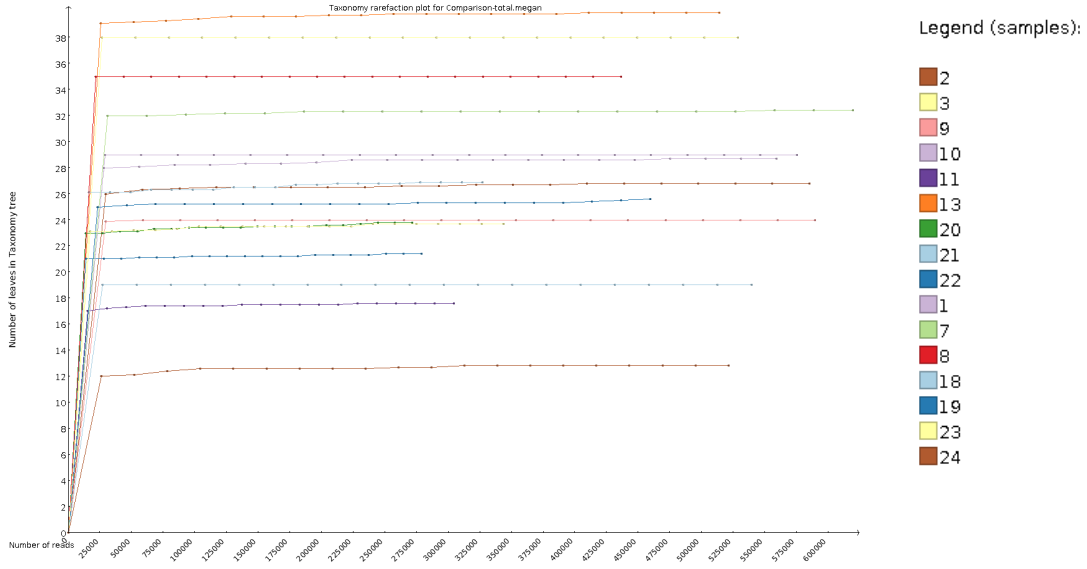


Figure 6. Taxonomy rarefaction plot for species richness in sixteen infants.

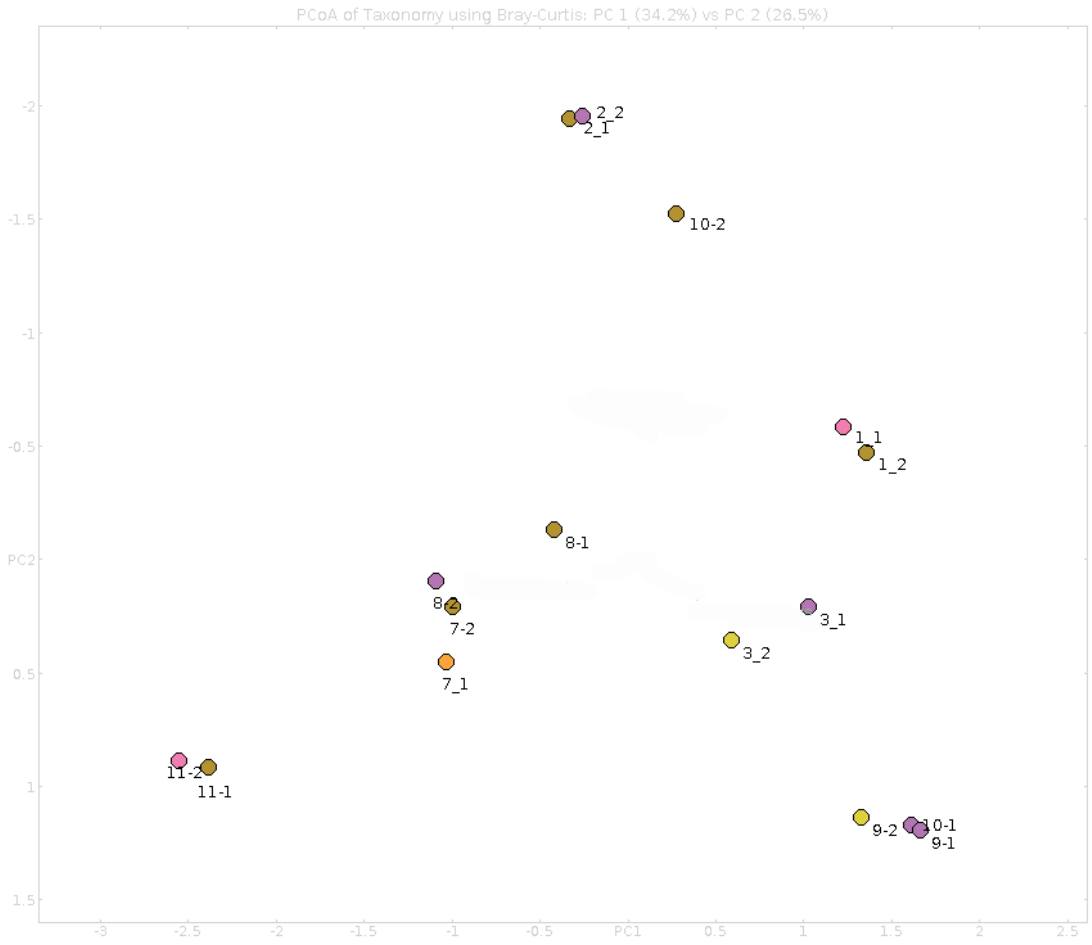
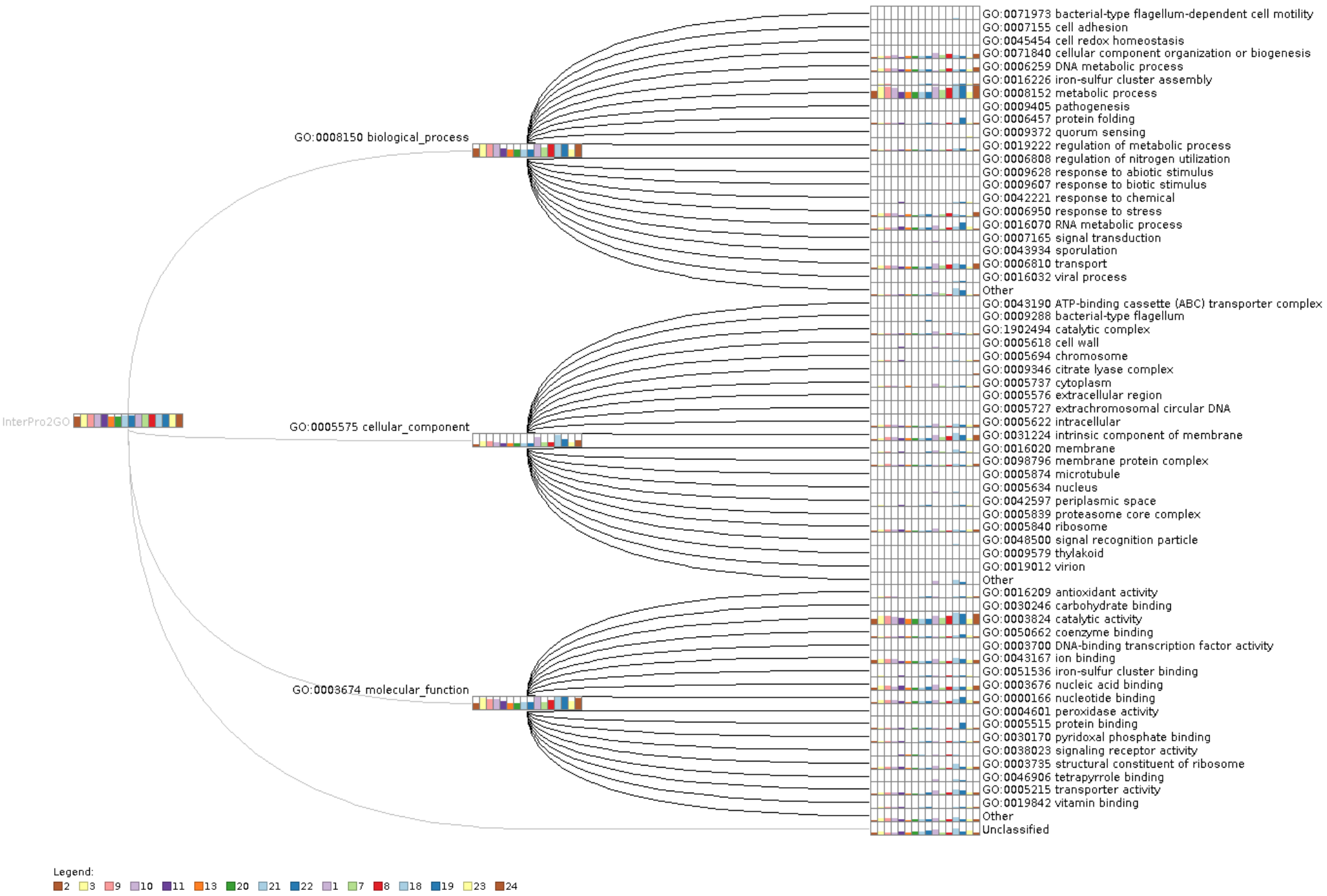


Figure 7. PCoA plot of eight infants (#2, #3, #9, #10 and #11 from breast milk group, #1, #7 and #8 from formula group) with separate two time points. Most of the beta diversity of each infant within two time points are more similar than across samples.

A)


InterPro2GO
GO:0008150 biological_process
GO:0005575 cellular_component
GO:0003674 molecular_function
GO:0071973 bacterial-type flagellum-dependent cell motility
GO:0007155 cell adhesion
GO:0045454 cell redox homeostasis
GO:0071840 cellular component organization or biogenesis
GO:0006259 DNA metabolic process
GO:0016226 iron-sulfur cluster assembly
GO:0008152 metabolic process
GO:0009405 pathogenesis
GO:0006457 protein folding
GO:0009372 quorum sensing
GO:0019222 regulation of metabolic process
GO:0006808 regulation of nitrogen utilization
GO:0009628 response to abiotic stimulus
GO:0009607 response to biotic stimulus
GO:0042221 response to chemical
GO:0006950 response to stress
GO:0016070 RNA metabolic process
GO:0007165 signal transduction
GO:0043934 sporulation
GO:0006810 transport
GO:0016032 viral process
Other
GO:0043190 ATP-binding cassette (ABC) transporter complex
GO:0009288 bacterial-type flagellum
GO:1902494 catalytic complex
GO:0005618 cell wall
GO:0005694 chromosome
GO:0009346 citrate lyase complex
GO:0005737 cytoplasm
GO:0005576 extracellular region
GO:0005727 extrachromosomal circular DNA
GO:0005622 intracellular
GO:0031224 intrinsic component of membrane
GO:0016020 membrane
GO:0098796 membrane protein complex
GO:0005874 microtubule
GO:0005634 nucleus
GO:0042597 periplasmic space
GO:0005839 proteasome core complex
GO:0005840 ribosome
GO:0048500 signal recognition particle
GO:0009579 thylakoid
GO:0019012 virion
Other
GO:0016209 antioxidant activity
GO:0030246 carbohydrate binding
GO:0003824 catalytic activity
GO:0050662 coenzyme binding
GO:0003700 DNA-binding transcription factor activity
GO:0043167 ion binding
GO:0051536 iron-sulfur cluster binding
GO:0003676 nucleic acid binding
GO:0000166 nucleotide binding
GO:0004601 peroxidase activity
GO:0005515 protein binding
GO:0030170 pyridoxal phosphate binding
GO:0038023 signaling receptor activity
GO:0003735 structural constituent of ribosome
GO:0046906 tetrapyrrole binding
GO:0005215 transporter activity
GO:0019842 vitamin binding
Other
Unclassified
Legend:
2 3 9 10 11 13 20 21 22 1 7 8 18 19 23 24

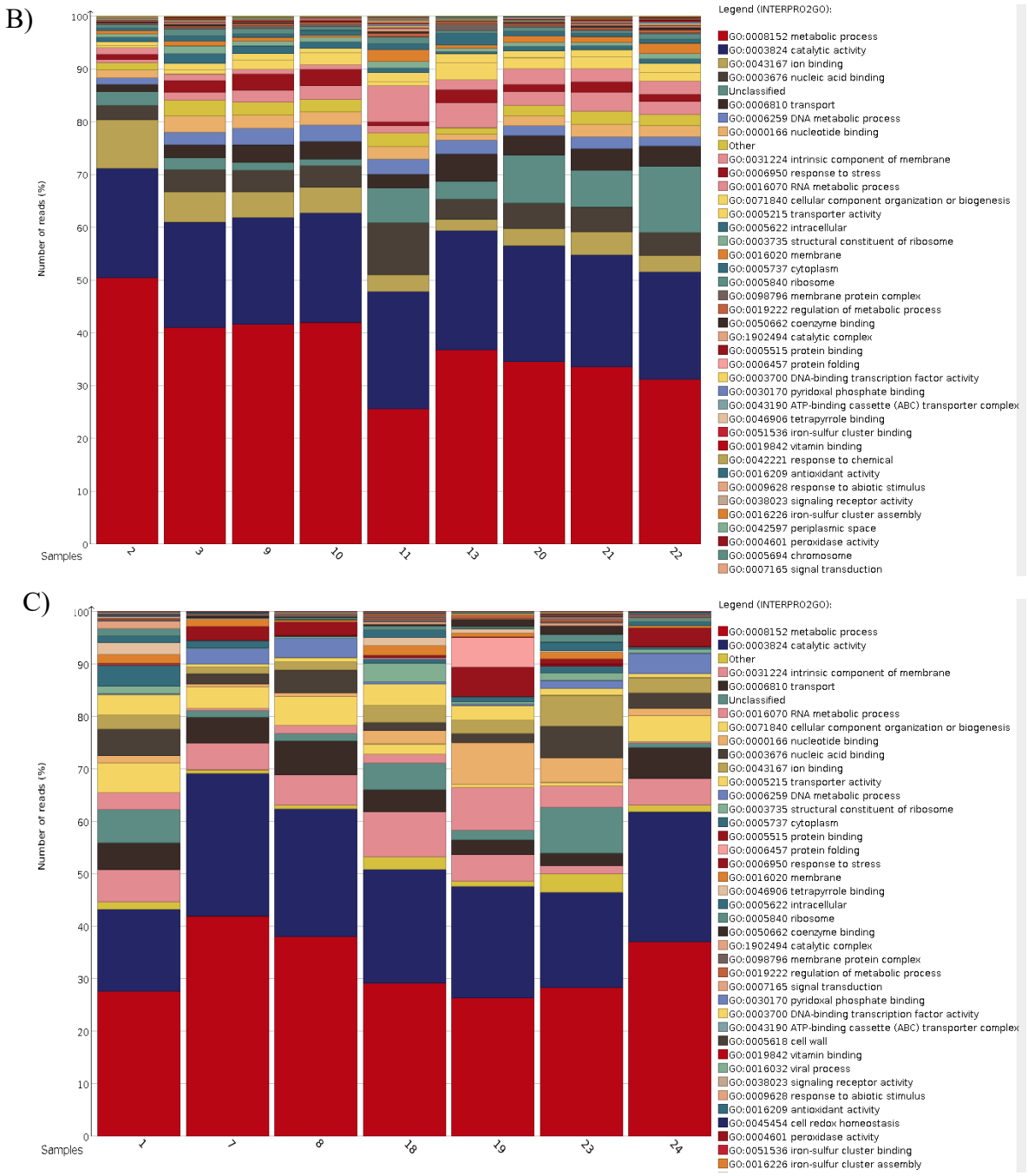


Figure 8. Functional analysis (InterPro2GO) of sixteen infants. (A) InterPro2GO tree of Metatranscriptome result of sixteen infants comparison. First nine columns are breast milk group and rest seven columns are formula group. (B) Bar chart of InterPro2GO in breast milk-fed infants. (Partial legend was showed) (C) Bar chart of InterPro2GO in formula-fed infants (Partial legend was showed).

4. DISCUSSION

In this study, we considered the effect of different diet and antibiotic exposure on preterm infants' intestinal flora. Regarding the alpha-diversity between different nutrition types, we achieved a conclusion similar to Gewolb et al. that there are no significant differences between breast milk-fed and formula-fed preterm infants in the first few days of their early life [76]. Since, in this study, most breast milk-fed babies also received minimal formula supplementation to meet nutrition needs, this may explain the similarity in alpha diversity between the two groups of infants. We also discovered through 16S rRNA sequencing that the bacteria composition of both breast milk and formula fed preterm infants at the phylum level are very similar and *Firmicutes* is the main predominate average at 68%, followed by *Proteobacteria* at 27%. On the genus level, we found that breast milk-fed preterm infants were mainly dominated by *Veillonella* and *Escherichia/Shigella*, followed by *Staphylococcus*, *Enterococcus*, *Clostridium*, *Streptococcus* and *Klebsiella*. In formula-fed infants, *Streptococcus* and *Klebsiella* are the main colonizers, and *Enterococcus*, *Staphylococcus*, *Veillonella* and *Clostridium* are the minor colonizers. These detected bacteria are all considered pathogenic bacteria or potentially pathogenic bacteria. For *Bifidobacterial,* which are considered beneficial bacteria, our results are similar to those of other intestinal studies in preterm infants [14, 47, 57, 76]; we also found low abundance of *Bifidobacteria* in the gut of preterm infants, which should have the most frequency and abundance in full-term healthy infants [23, 24, 48]. The *Propionibacterium*, the only bacteria species found in this study to have significant differences between different nutrition sources, was also found by a previous study in the gut of infants who were born through C-section, and it was considered to have been

obtained from the mother's skin during the C-section process [59]. However, *Propionibacterium* lives in and around the human sebaceous glands and sweat glands and this bacterium was confirmed common bacterium in breast milk [85-87]. Therefore, we believe that in our study breast milk-fed infants most likely received *Propionibacterium* from the milk or direct contact through the process of breastfeeding. In Table 3, both OTU 34 and OTU 77 correspond to *Nitrososphaera*, which belongs to the *Thaumarchaeota* phylum Archaea. *Nitrososphaera* was found in the adult human gut and considered to be related to diet and other microbes [88]. Hoffmann at el. observed that *Nitrososphaera* has a positive association with protein ingestion [88]. In this study, *Nitrososphaera* has more frequency in formula-fed preterm infants. We speculate that proteins in formula milk promote the growth of *Nitrososphaera*. OTU4, which corresponds to *Staphylococcus*, was detected more abundantly in breast milk-fed infants. Since *Staphylococcus* is a common bacteria in breast milk [49], this result is what we expected. *Staphylococcus aureus*, one of the common pathogen species of *Staphylococcus*, has been confirmed to be primarily responsible for mastitis on lactating women [89]. Some studies have shown that *Staphylococcus aureus* can be transmitted to preterm infants through breast milk and affect infants intestinal colonization [90, 91]. *Group B Streptococcus* (GBS) is responsible for common neonatal diseases, such as neonatal sepsis [92, 93] and meningitis [94], but the transmission mechanism from breast milk to infants has not been confirmed. Not only that, Doare et al. reported that breast milk may contain specific antibodies that could inhibit *Streptococcus* in infants [95]. In this study, although OTU 37, which represent *Streptococcus*, has more frequency in breast milk-fed infants, the sintax results show that *Streptococcus* was more abundant in the formula

group than breast milk group, at 18.6% and 9.3% respectively. *Rothia* is a normal bacteria community in the oral or respiratory tract [96], so it is not surprising that it is also found in the intestines. Since it is not a major colonizer in infant's intestines, there are few studies focused on it. *Rothia* was found more abundantly in breast milk-fed infant's fecal samples by Wang et al. [97] and to exist in breast milk by Jost et al. [98]. However, Biesbroek et al. reported greater abundance of *Rothia* in formula-fed infants gut [99], which is consistent with our findings.

The metatranscriptome results show that breast milk-fed preterm infants have more abundance of *Streptococcus*, *Clostridium*, *Escherichia* and *Veillonella* than formula-fed; and formula-fed infants contain more *Enterococcus, Pantoea* and *Klebsiella*. *Bifidobacterium* and *Staphylococcus* exist in both groups and the difference in abundance could not be observed. Most of the differences between the breast milk-fed group and formula-fed group are consistent with the 16S, but an inconsistent result was found with *Streptococcus*.

Metatranscriptome is considered to be more unbiased. Unlike 16S rRNA sequencing, metatranscriptome does not undergo a PCR step, which could cause some uneven amplification. Our results confirm this in that the proportion of each bacteria in metatranscriptome results and 16S results are different. In addition, the metatranscriptome data can be analyzed for gene function and provide a better understanding of microbiome. However, because it requires more precise operations and more expensive costs, metatranscriptome is not as widely used as 16S in microbiome studies.

There are still many limitations in our study, including small sample size, two time points too close, varied times of the antibiotics use as well as different delivery methods and gender etc., which may complicate the data analysis. Nevertheless, our results still provide new knowledge on the understanding of the effect of diet on intestinal microbial composition of moderate to late premature infants.

CHAPTER 3:

GUT VIROME IN THE BREAST MILK- AND FORMULA-FED PRETERM INFANTS GTU

1. INTRODUCTION

Compared to bacteria, virus abundance in the gut of the early stage of infants is very low. Because of this, we know little about how the virus colonizes in infants and how it can be affected by different factors. Bacteriophage of *Caudovirales* order and DNA virus of *Anelloviruses* were reported multiple times by different studies that exist in early life of infant gut[13, 64, 66]. Lim et al. also reports that the abundance of bacteriophage will shift from *Caudovirales* to *Microviridae* through first year of their life.

Unlike bacteria, the effects of diet on the virome are still ambiguous. Pannaraj et al. claimed there are 30% of viral contigs were shared by mothers' breast milk and infants' fecal samples[68]. However, some other studies believe that the most abundant viruses in infant gut were not directedly obtained from diet[64, 69]. Additionally, the antibiotic influencing on early life virome are not clear. We are trying to understand the patterns and influencing factors of early colonization of the virus in preterm infants.

2. MATERIAL AND METHODS

2.1. Sample collection.

All the experiments that involved human subjects in this study were approved by Sanford Health IRA in Sioux Falls, South Dakota. Total 20 preterm infants who were born between 32 0/7 to 36 6/7 weeks gestation were recruited by us and fecal specimens were collected in Sanford Children's MB2 Clinic, Sioux Falls, during March 30 2017 to December 10 2017. The bioinformation of infants and mothers are described in **Table 1**.

Ten of twenty were fed by breast milk and ten were fed by formula milk. Eleven infants were received Ampicillin/Gentamicin antibiotics as treatments. Six of twenty were born through vaginal delivery and fourteen were born via Cesarean section. Additionally, there are five pairs of twins: infant #7 and #8; #9 and #10; #18 and #19; #21 and #22; #23 and #24. For each infant, we had two time points for sample collection.

Two different methods of storing samples were used. The fecal sample collected from each infant was separately stored into DNA/RNA shield tubes (Zymo Research) keep in 4 ºC and sterile 15ml centrifuge tubes keep in -80 ºC.

2.2. RNA extraction, library preparation and RNA sequencing

Fresh samples that stored in sterile tubes were suspended with five times volume of phosphate-buffered saline (PBS). Samples from the two time points were combined into one. After centrifugation at 5000rpm for 5min, the supernatants were filtered the supernatant with 0.45 μm filters. DNA/RNA were extracted by using DNA/RNA Mini Kit (ZRC188678, ZymoBIOMICSTM). The concentrations of DNA/ RNA were measured with Nanodrop 2000. A total of 12 μl RNA was used to deplete rRNAs using rRNA Depletion Kit (E6310L, NEBNext®) by following the manufacturer's instructions. Library preparation was performed by using Directional RNA Library Prep Kit for Illumina kit (E7760S, NEBNext®).

2.3. Metagenomic analysis.

The Metagenomic data was trimmed by Trimmomatic[80] first to cut the adaptors off and then Trinity[100] was used for assembly. The Trinity outputs were run through the blastx by using USEARCH. Extracted the blastx results that contained "phage" in description, and were double-checked with BLAST command, then generate into a fasta

file. Bowtie was used to build index of the fasta file, and the counts of reads were obtained through bowtie_counts command.

2.4. Transmission Electron Microscopy (TEM) test.

The fresh samples that stored in sterile tubes were suspended with five times volume of phosphate-buffered saline (PBS). After vortex to loose the stool samples and break up the big chucks, samples were centrifuged at 8,000 rpm for 5 min and the supernatants were filtered with 0.45 μm filters. The filtrate was transferred into ultracentrifuge tubes and balanced with PBS. Ultracentrifugation was performed in a Beckman Coulter Optima MAX Ultracentrifuge at 50,000 rpm for 90 min. Supernatants were discarded and the pellets were resuspended in PBS and UV-inactivated. The samples were sent to the Electron Microscope Shared Resource Laboratory at University of Rochester Medical Center for sample fixation and TEM imaging.

3. RESULTS

3.1. Metatranscriptome

We found 29 potential phage contigs in all preterm infants' fecal samples after comparison with NCBI nr database (Table 5). Those 29 contigs may come from 19 different phage species. Most of the virus counts are more in the formula fed group than breast milk fed except 5 potential phage contigs. Most of them are considered as dsDNA viruses and belongs to family *Myoviridae*, *Podoviridae* or *Siphoviridae*. Uncultured Mediterranean phage sequences were identified.

For the three different bacteriophage families, we compared their read counts between two different nutrition groups (Figure 9). We found that the read count of *Siphoviridae* are significantly higher in formula-fed than breast milk-fed preterm infants.

Table 5. Summary of potential phage contigs results. Classifier against with database from NCBI.

Name	Contig length (bp)	Description	Sequence ID	Identity	Genotype	Virus family	Read counts of Nutrition *	
							B	F
Myoviridae like phage	349	hypothetical protein (Myoviridae sp.)	AXH72494.1	40%	dsDNA viruses	Myoviridae	2	10
Salicola phage SCTP-2 like phage	223	hypothetical protein PBI_SCTP2_459 (Salicola phage SCTP-2)	ASV44474.1	59%	dsDNA viruses	Myoviridae	0	5
Cronobacter phage vB_CsaP_Ss1 like phage	275	hypothetical protein SS1_30 (Cronobacter phage vB_CsaP_Ss1)	AIK67534.1	58%	dsDNA viruses	Podoviridae	1	4
Staphylococcus phage St 134 like phage	220		AQT25391.1	69%	dsDNA viruses	Podoviridae	0	6
Arthrobacter phage Amigo like phage 1	385	DNA primase (Arthrobacter phage Amigo)	ALY08407.1	57%	dsDNA viruses	Siphoviridae	1	9
Arthrobacter phage Amigo like phage 2	311	major capsid subunit (Arthrobacter phage Amigo)	ALY08384.1	40%	dsDNA viruses	Siphoviridae	5	6
Arthrobacter phage Molivia like phage	298	DNA helicase (Arthrobacter phage Molivia)	ASX99295.1	48%	dsDNA viruses	Siphoviridae	4	9
Arthrobacter phage Rings like phage 1	205	terminase large subunit (Arthrobacter phage Rings)	ALY10098.1	55%	dsDNA viruses	Siphoviridae	0	9
Arthrobacter phage Rings like phage 2	383	terminase large subunit (Arthrobacter phage Rings)	ALY10098.1	48%	dsDNA viruses	Siphoviridae	5	11
Clavibacter phage CMP1 like phage 1	243	hypothetical protein CMP1-12 (Clavibacter phage CMP1)	YP_003359103.1	41%	dsDNA viruses	Siphoviridae	7	3
Clavibacter phage CMP1 like phage 2	295	hypothetical protein CMP1-31 (Clavibacter phage CMP1)	YP_003359122.1	44%	dsDNA viruses	Siphoviridae	3	5
Clavibacter phage CMP1 like phage 3	302	hypothetical protein CMP1-44 (Clavibacter phage CMP1)	YP_003359135.1	36%	dsDNA viruses	Siphoviridae	1	14
Clavibacter phage CN1A like phage 1	227	helicase (Clavibacter phage CN1A)	YP_009004226.1	45%	dsDNA viruses	Siphoviridae	3	10
Clavibacter phage CN1A like phage 2	267	terminase large subunit (Clavibacter phage CN1A)	YP_009004271.1	42%	dsDNA viruses	Siphoviridae	4	8
Erysipelothrix phage phi1605 like phage	203	DNA helicase (Erysipelothrix phage phi1605)	ASD51083.1	60%	dsDNA viruses	Siphoviridae	1	3
Gordonia phage Fury like phage	616	RuvC-like resolvase (Gordonia phage Fury)	AXH49791.1	43%	dsDNA viruses	Siphoviridae	5	21
Mycobacterium phage SiSi like phage	303	WhiB (Mycobacterium phage SiSi)	YP_008051183.1	42%	dsDNA viruses	Siphoviridae	3	6
Streptomyces phage Ibantik like phage	235	DnaQ-like exonuclease (Streptomyces phage Ibantik)	AWN05257.1	55%	dsDNA viruses	Siphoviridae	4	4
Streptomyces phage Satis like phage	265	hypothetical protein SEA_SATIS_154 (Streptomyces phage Satis)	AXH66313.1	81%	dsDNA viruses	Siphoviridae	3	3
Methylophilales phage HIM624-A like phage	391	hypothetical protein MTPG_00031 (Methylophilales phage HIM624-A)	AFB70782.1	56%	dsDNA viruses	unclassified dsDNA phages	10	7
uncultured Mediterranean phage uvDeep1-CGR2-KM23-C896 like phage 1	536	hypothetical protein (uncultured Mediterranean phage uvDeep1-CGR2-KM23-C896)	ANS03047.1	77%	N/A	uncultured environmental isolates	4	13
uncultured Mediterranean phage uvDeep1-CGR2-KM23-C896 like phage 2	286	hypothetical protein (uncultured Mediterranean phage uvDeep1-CGR2-KM23-C896)	ANS03047.1	63%	N/A	uncultured environmental isolates	3	8
uncultured Mediterranean phage uvDeep-CGR1-KM17-C101 like phage	259	hypothetical protein (uncultured Mediterranean phage uvDeep-CGR1-KM17-C101)	ANS02997.1	76%	N/A	uncultured environmental isolates	3	6
uncultured Mediterranean phage uvMED like phage 1	239	DNA polymerase elongation subunit (family B) (PolB) (uncultured Mediterranean phage uvMED)	BAR35078.1	63%	N/A	environmental samples	4	3
uncultured Mediterranean phage uvMED like phage 2	247	hypothetical protein (uncultured Mediterranean phage uvMED)	BAQ92863.1	58%	N/A	environmental samples	3	7
uncultured Mediterranean phage uvMED like phage 3	224	phage terminase, large subunit, PBSX family (TIGR01547) (uncultured Mediterranean phage uvMED)	BAR33829.1	58%	N/A	environmental samples	3	3
uncultured Mediterranean phage uvMED like phage 4	259	tail fiber protein (uncultured Mediterranean phage uvMED)	BAQ93890.1	49%	N/A	environmental samples	2	7
uncultured Mediterranean phage like phage 1	370	cytidylyltransferase like protein (uncultured Mediterranean phage)	ANS05297.1	47%	N/A	environmental samples	5	3
uncultured Mediterranean phage like phage 2	453	hypothetical protein (uncultured Mediterranean phage)	ANS05305.1	67%	N/A	environmental samples	1	10

* In nutrition groups, B = Breast milk fed preterm infants, F = Formula fed preterm infants

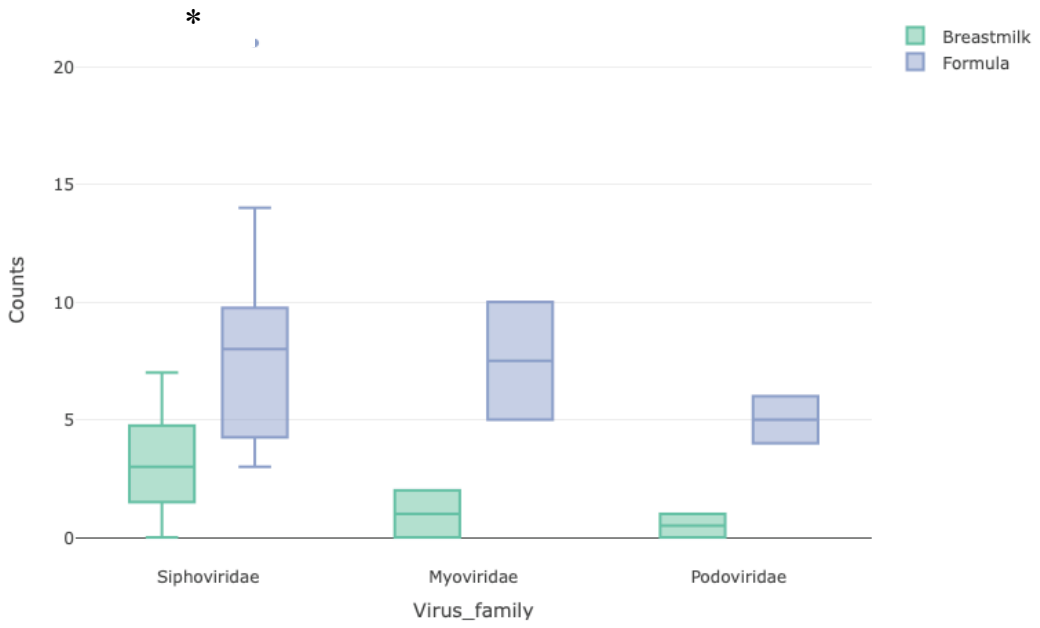


Figure 9. Boxplot of read counts in three detected viruses between different nutrition. * Read counts of *Siphoviridae* have significant difference between different nutrition. p-value = 0.002158

3.2. Transmission Electron Microscopy (TEM) results

We randomly chose four samples from the breast milk and formula fed infants for TEM examination. Infant #3 and #11 were fed with breast milk. Infant #25 and #27 were formula fed. Numerous *Siphoviridae* liked phage particles were observed in infant #25, but not in the other three infants (Figure 13). No other phage-like particles were observed in these four samples. Some spherical virus-like particles were also observed in both formula-fed infants. Virus-like structures were not observed in the two breast milk-fed infant samples.

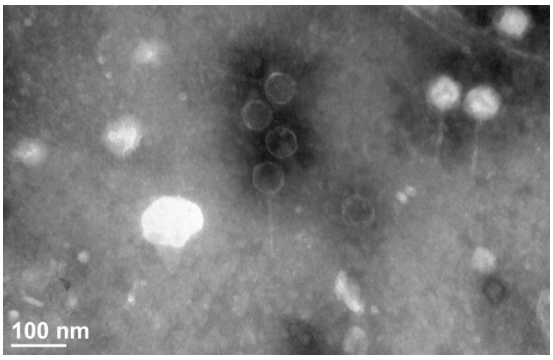


Figure 10. Numerous *Siphoviridae* liked phage particles were observed in sample #25 under TEM.

4. DISCUSSION

Siphoviridae, *Myoviridae* and *Podoviridae* are the most abundant bacteriophages detected in the late preterm infants. Our results are consistent with the conclusions of Lim et al.[13] and Breitbart et al.[64]. It appears that formula-fed infants showed significantly higher *Siphoviridae* read counts than the breast milk-fed infants. Additionally, we observed numerous *Siphoviridae*-like phage particles in one the formula-fed infant. Similarly, antibiotic use infants seemed to show significantly higher read counts of *Siphoviridae*. The significance of these observations remains to be validated by future studies due to our small sample size. Additionally, since we did not purify virus-like particles as typically used by other studies on gut virome (ref), our RNAseq libraries contain abundant bacterial sequences, which may contribute to the relatively small numbers of viruses.

It is regrettable that some factors constrain our experimental choices. The main factor is the amount of fresh samples are too low that did not allow us to do ultracentrifuge before filter. This has led to the existence of a large number of bacteria in RNAseq library, diluted the proportion of the viruses. In order to better understand the

intestinal colonization of premature infants, more follow-up experiments are needed. One direction is that we can further isolate the suspected *Siphoviridae* phage seen on the TEM, segueing to identify the species and try to figure out the host range, to understand the interaction between phage and bacteria in preterm infant gut. We can also try to find more viruses by using PCR amplification technique with primers of different viruses, especially the common viruses in human gut such as crAssphage, *Anelloviruses*, and the bacteriophages that belongs to *Microviridae* family[13, 64, 65].

APPENDIX

Supplementary Table 1. Alpha diversities in different nutrition.

Index				p-value
berger_parker	Fomular	~	Breast milk	0.48
buzas_gibson	Fomular	~	Breast milk	0.684
chao1	Fomular	~	Breast milk	0.794
dominance	Fomular	<	Breast milk	0.142
equitability	Fomular	<	Breast milk	0.123
jost	Fomular	<	Breast milk	0.124
jost1	Fomular	<	Breast milk	0.141
reads	Fomular	<	Breast milk	0.142
richness	Fomular	~	Breast milk	0.483
richness2	Fomular	~	Breast milk	0.315
robbins	Fomular	<	Breast milk	0.164
simpson	Fomular	>	Breast milk	0.142
shannon_e	Fomular	<	Breast milk	0.142
shannon_2	Fomular	<	Breast milk	0.143
shannon_10	Fomular	<	Breast milk	0.143
flyvbjerg	Fomular	~	Breast milk	0.529
mirror	Fomular	<	Breast milk	0.19
mirrorns	Fomular	~	Breast milk	0.852
logfit	Fomular	>	Breast milk	0.137
logfitns	Fomular	~	Breast milk	0.222
logfitmu	Fomular	>	Breast milk	0.192
logfitmuns	Fomular	~	Breast milk	0.337

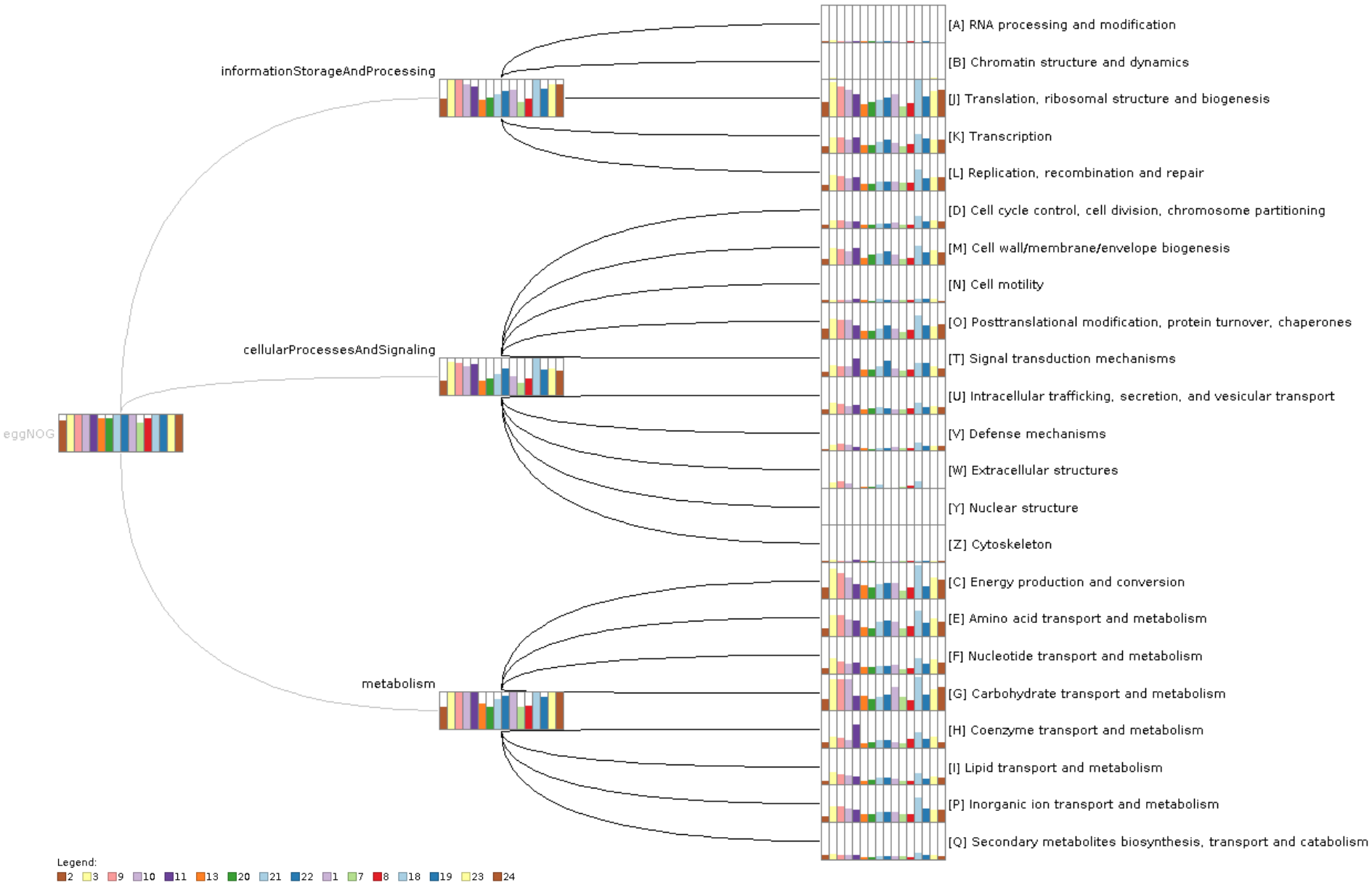
"~" and "=" = metric is approximately equal in both groups
">" and "<" = metric has weak significance ($P < 0.2$)
">>" = metric has high significant ($P < 0.05$)

Supplementary Table 2. Alpha diversities in antibiotic using.

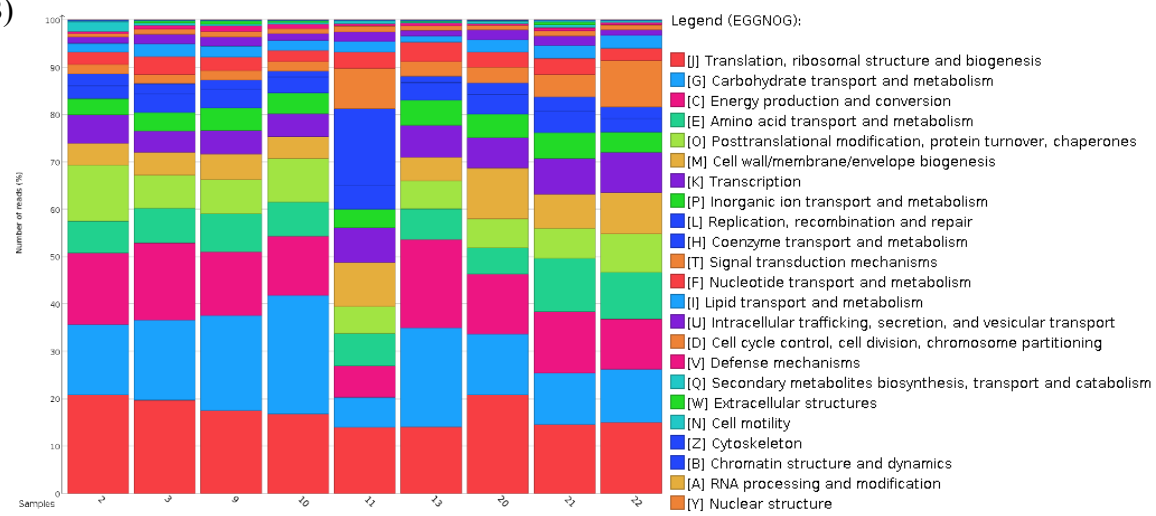
Index				p-value
berger_parker	N	~	A	0.202
buzas_gibson	N	>>	A	0.00789
chao1	N	~	A	0.941
dominance	N	~	A	0.259
equitability	N	~	A	0.293
jost	N	~	A	0.26
jost1	N	~	A	0.228
reads	N	~	A	0.458
richness	N	~	A	0.824
richness2	N	~	A	0.766
robbins	N	~	A	0.33
simpson	N	~	A	0.262
shannon_e	N	~	A	0.23
shannon_2	N	~	A	0.232
shannon_10	N	~	A	0.229
flyvbjerg	N	~	A	0.553
mirror	N	~	A	1
mirrorns	N	~	A	0.603
logfit	N	=	A	0.84
logfitns	N	~	A	0.589
logfitmu	N	~	A	0.299
logfitmuns	N	~	A	0.398

N = Non-antibiotic using preterm infants
A = Antibiotic using preterm infants
"~" and "=" = metric is approximately equal in both groups
">" and "<" = metric has weak significance ($P < 0.2$)
">>" = metric has high significant ($P < 0.05$)

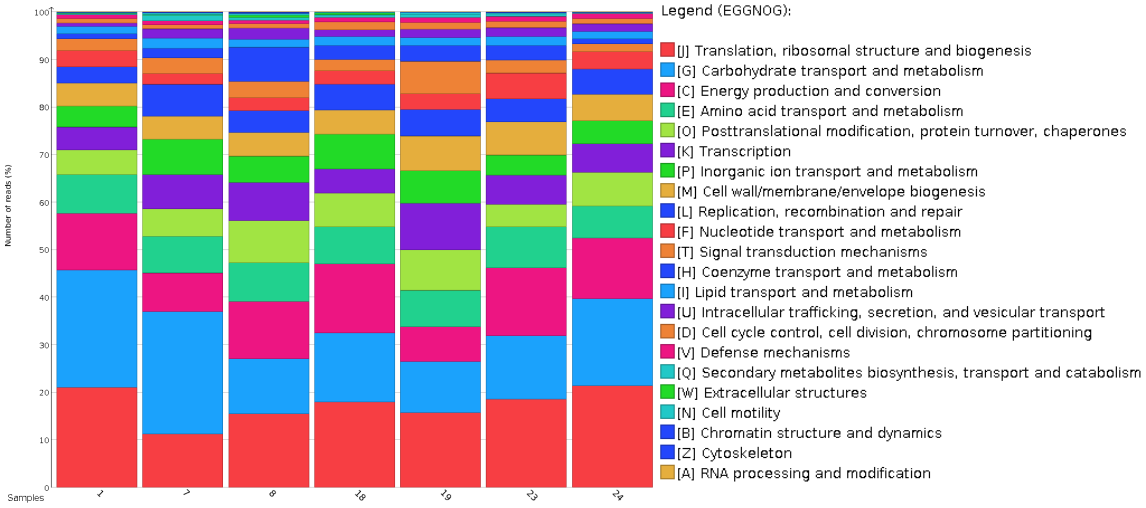
A)


eggNOG
informationStorageAndProcessing
cellularProcessesAndSignaling
metabolism
[A] RNA processing and modification
[B] Chromatin structure and dynamics
[J] Translation, ribosomal structure and biogenesis
[K] Transcription
[L] Replication, recombination and repair
[D] Cell cycle control, cell division, chromosome partitioning
[M] Cell wall/membrane/envelope biogenesis
[N] Cell motility
[O] Posttranslational modification, protein turnover, chaperones
[T] Signal transduction mechanisms
[U] Intracellular trafficking, secretion, and vesicular transport
[V] Defense mechanisms
[W] Extracellular structures
[Y] Nuclear structure
[Z] Cytoskeleton
[C] Energy production and conversion
[E] Amino acid transport and metabolism
[F] Nucleotide transport and metabolism
[G] Carbohydrate transport and metabolism
[H] Coenzyme transport and metabolism
[I] Lipid transport and metabolism
[P] Inorganic ion transport and metabolism
[Q] Secondary metabolites biosynthesis, transport and catabolism
Legend:
2 3 9 10 11 13 20 21 22 1 7 8 18 19 23 24

B)

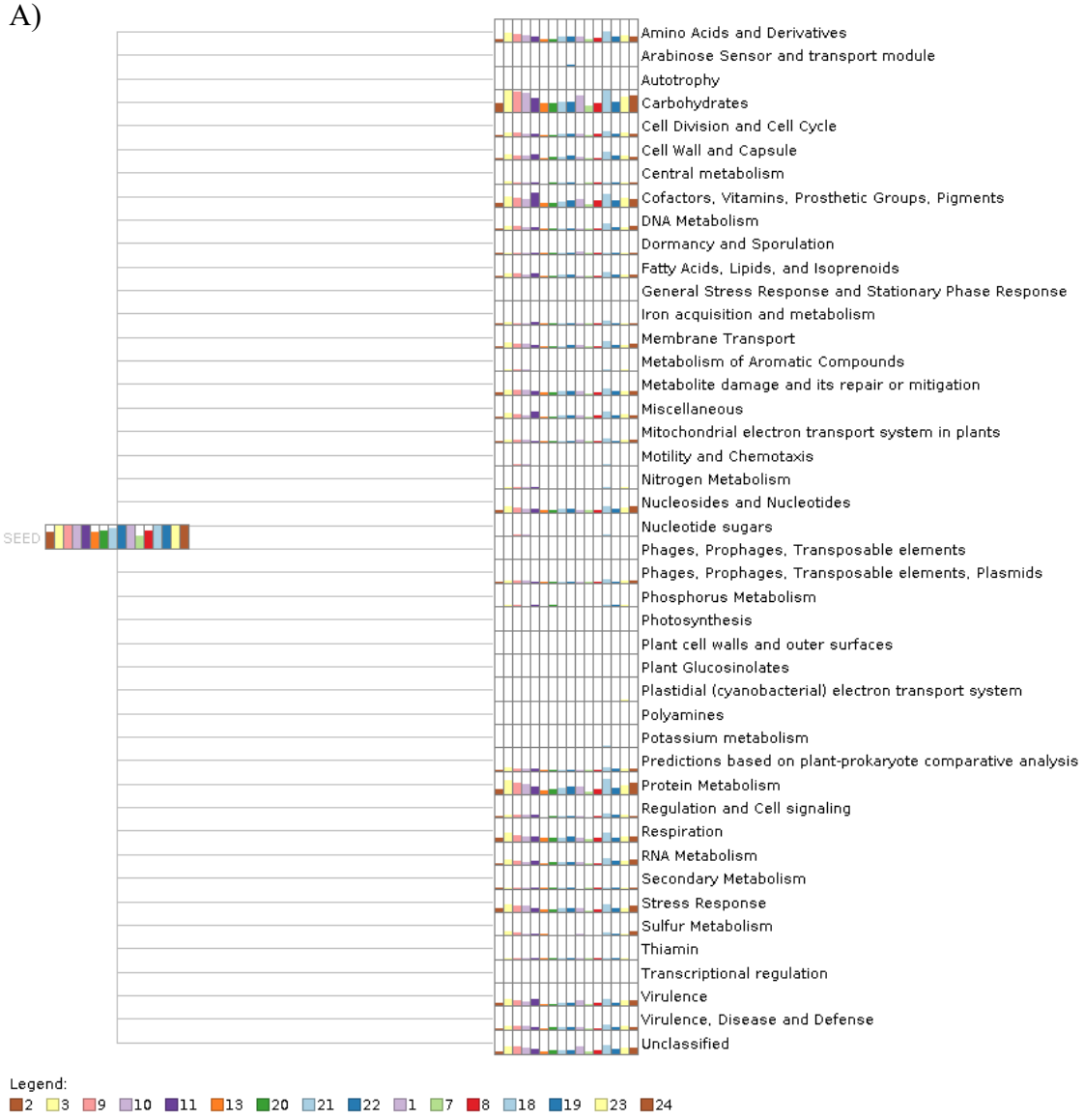


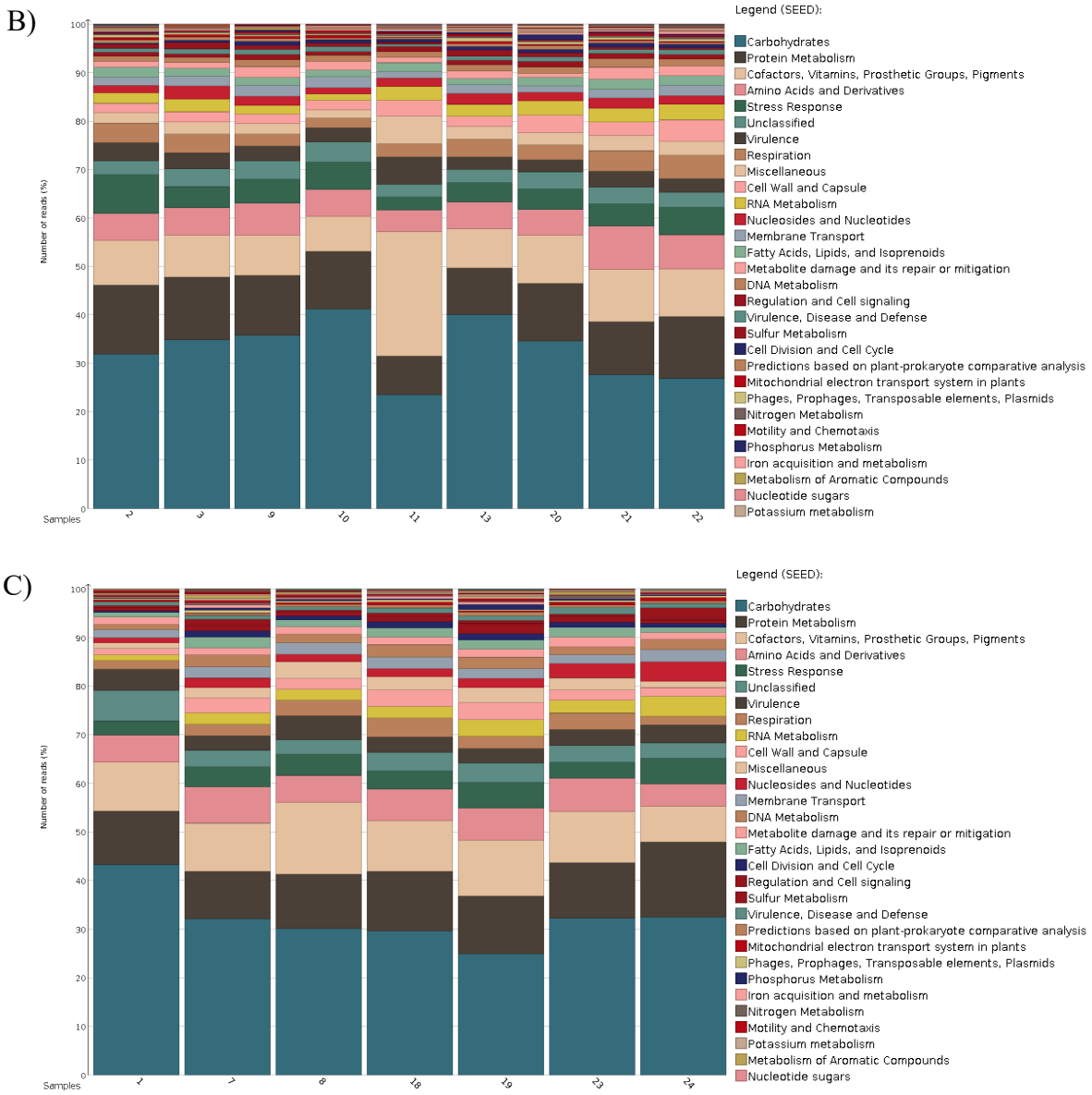
C)



Supplementary figure 1. Functional analysis (eggNOG) of sixteen infants. (A) EggNOG tree of Metatranscriptome result of sixteen infants comparison. (B) Bar chart of EggNOG in breast milk-fed infants (Partial legend was showed). (C) Bar chart of EggNOG in formula-fed infants (Partial legend was showed).

A)


Amino Acids and Derivatives
Arabinose Sensor and transport module
Autotrophy
Carbohydrates
Cell Division and Cell Cycle
Cell Wall and Capsule
Central metabolism
Cofactors, Vitamins, Prosthetic Groups, Pigments
DNA Metabolism
Dormancy and Sporulation
Fatty Acids, Lipids, and Isoprenoids
General Stress Response and Stationary Phase Response
Iron acquisition and metabolism
Membrane Transport
Metabolism of Aromatic Compounds
Metabolite damage and its repair or mitigation
Miscellaneous
Mitochondrial electron transport system in plants
Motility and Chemotaxis
Nitrogen Metabolism
Nucleosides and Nucleotides
Nucleotide sugars
Phages, Prophages, Transposable elements
Phages, Prophages, Transposable elements, Plasmids
Phosphorus Metabolism
Photosynthesis
Plant cell walls and outer surfaces
Plant Glucosinolates
Plastidial (cyanobacterial) electron transport system
Polyamines
Potassium metabolism
Predictions based on plant-prokaryote comparative analysis
Protein Metabolism
Regulation and Cell signaling
Respiration
RNA Metabolism
Secondary Metabolism
Stress Response
Sulfur Metabolism
Thiamin
Transcriptional regulation
Virulence
Virulence, Disease and Defense
Unclassified
SEED
Legend:
2 3 9 10 11 13 20 21 22 1 7 8 18 19 23 24



Supplementary figure 2. Functional analysis (SEED) of sixteen infants. (A) SEED tree of Metatranscriptome result of sixteen infants comparison. (B) Bar chart of SEED in breast milk-fed infants (Partial legend was showed). (C) Bar chart of SEED in formula-fed infants (Partial legend was showed).

LITERATURE CITED

1. Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
2. D'Argenio, V. and F. Salvatore, *The role of the gut microbiome in the healthy adult status.* Clinica Chimica Acta, 2015. **451**: p. 97-102.
3. Gordon, P.J.T.a.J.I., *The core gut microbiome, energy balance and obesity.* J Physiol, 2009. **587**(17): p. 4153-4158.
4. Tap, J., et al., *Towards the human intestinal microbiota phylogenetic core.* Environmental microbiology, 2009. **11**(10): p. 2574-2584.
5. Lozupone, C.A., et al., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220.
6. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* nature, 2010. **464**(7285): p. 59.
7. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora.* science, 2005. **308**(5728): p. 1635-1638.
8. Abu-Shanab, A. and E.M. Quigley, *The role of the gut microbiota in nonalcoholic fatty liver disease.* Nature reviews Gastroenterology & hepatology, 2010. **7**(12): p. 691.
9. Cho, I. and M.J. Blaser, *The human microbiome: at the interface of health and disease.* Nature Reviews Genetics, 2012. **13**(4): p. 260.
10. Ley, R.E., et al., *Microbial ecology: human gut microbes associated with obesity.* nature, 2006. **444**(7122): p. 1022.
11. Flint, H.J., et al., *The role of the gut microbiota in nutrition and health.* Nature reviews Gastroenterology & hepatology, 2012. **9**(10): p. 577.
12. Tang, W.W., et al., *Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk.* New England Journal of Medicine, 2013. **368**(17): p. 1575-1584.
13. Lim, E.S., et al., *Early life dynamics of the human gut virome and bacterial microbiome in infants.* Nature medicine, 2015. **21**(10): p. 1228.
14. Palmer, C., et al., *Development of the Human Infant Intestinal Microbiota.* PLoS Biology, 2007. **5**(7): p. 1556-1573.
15. Neu, J., *Preterm infant nutrition, gut bacteria, and necrotizing enterocolitis.* Curr Opin Clin Nutr Metab Care., 2015. **18**(3): p. 285-288.
16. Lim, E.S., D. Wang, and L.R. Holtz, *The Bacterial Microbiome and Virome Milestones of Infant Development.* Trends in Microbiology, 2016. **24**(10): p. 801-810.
17. Li, M., M. Wang, and S.M. Donovan. *Early development of the gut microbiome and immune-mediated childhood disorders*. in *Seminars in reproductive medicine*. 2014. Thieme Medical Publishers.
18. Mackie, R.I., A. Sghir, and H.R. Gaskins, *Developmental microbial ecology of the neonatal gastrointestinal tract.* Am J Clin Nutr, 1999(69): p. 1035S–1045S.
19. Jiménez, E., et al., *Is meconium from healthy newborns actually sterile?* Res Microbiol, 2008. **159**(3): p. 187-193.
20. Jiménez, E., et al., *Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section.* Current microbiology, 2005. **51**(4): p. 270-274.

21. Bearfield, C., Davenport, E.S., Sivapathasundaram, V., Allaker, R.P., *Possible association between amniotic fluid micro-organism in- fection and microflora in the mouth.* British Journal of Obstetrics and Gynaecology, 2002. **109**: p. 527-533.
22. Wagner, C.L., S.N. Taylor, and D. Johnson, *Host factors in amniotic fluid and breast milk that contribute to gut maturation.* Clinical reviews in allergy & immunology, 2008. **34**(2): p. 191-204.
23. Benno, Y., et al., *The intestinal microflora of infants: composition of fecal flora in breast‐fed and bottle‐fed infants.* 1984. **28**(9): p. 975-986.
24. Stark, P.L. and A. Lee, *The microbial ecology of the large bowel of breastfed and formula-fed infants during the first year of life.* Journal of medical microbiology, 1982. **15**(2): p. 189-203.
25. Hall, M., et al., *Factors influencing the presence of faecal lactobacilli in early infancy.* 1990. **65**(2): p. 185-188.
26. Stark, P. and A. Lee, *The bacterial colonization of the large bowel of pre-term low birth weight neonates.* Epidemiology & Infection, 1982. **89**(1): p. 59-67.
27. Blakey, J.L., et al., *Development of gut colonization in pre-term neonates.* Journal of medical microbiology, 1982. **15**(4): p. 519-529.
28. Magne, F., et al., *Low species diversity and high interindividual variability in faeces of preterm infants as revealed by sequences of 16S rRNA genes and PCR-temporal temperature gradient gel electrophoresis profiles.* 2006. **57**(1): p. 128-138.
29. Sakata, H., H. Yoshioka, and K. Fujita, *Development of the intestinal flora in very low birth weight infants compared to normal full-term newborns.* Eur J Pediatr, 1985. **144**(2): p. 186-190.
30. Westerbeek, E.A., et al., *The intestinal bacterial colonisation in preterm infants: a review of the literature.* Clinical nutrition, 2006. **25**(3): p. 361-368.
31. Rotimi, V., et al., *The development of bacterial flora of premature neonates.* 1985. **94**(3): p. 309-318.
32. Magne, F., et al., *Fecal microbial community in preterm infants.* Journal of pediatric gastroenterology and nutrition, 2005. **41**(4): p. 386-392.
33. Bäckhed, F., et al., *Dynamics and stabilization of the human gut microbiome during the first year of life.* Cell host & microbe, 2015. **17**(5): p. 690-703.
34. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* 2012. **486**(7402): p. 222.
35. Quigley, M. and W.J.C.D.o.s.r. McGuire, *Formula versus donor breast milk for feeding preterm or low birth weight infants.* 2014(4).
36. Gross, S.J., *Growth and biochemical response of preterm infants fed human milk or modified infant formula.* New England Journal of Medicine, 1983. **308**(5): p. 237-241.
37. Oddy, W.H., *The impact of breastmilk on infant and child health.* Breastfeeding Review, 2002. **10**(3): p. 5.
38. Hanson, L., et al., *New knowledge in human milk immunoglobulin.* Acta Pædiatrica, 1978. **67**(5): p. 577-582.
39. Hamosh, M., *Bioactive factors in human milk.* Pediatric Clinics of North America, 2001. **48**(1): p. 69-86.

40. Peterson, J.A., S. Patton, and M. Hamosh, *Glycoproteins of the human milk fat globule in the protection of the breast-fed infant against infections.* Neonatology, 1998. **74**(2): p. 143-162.
41. Wold, A. and I. Adlerberth, *Breast feeding and the intestinal microflora of the infant—implications for protection against infectious diseases*, in *Short and long term effects of breast feeding on child health*. 2002, Springer. p. 77-93.
42. Lucas, A. and T.J.J.T.L. Cole, *Breast milk and neonatal necrotising enterocolitis.* 1990. **336**(8730-8731): p. 1519-1523.
43. Boyd, C.A., M.A. Quigley, and P. Brocklehurst, *Donor breast milk versus infant formula for preterm infants: systematic review and meta-analysis.* Archives of Disease in Childhood-Fetal and Neonatal Edition, 2007. **92**(3): p. F169-F175.
44. CLAUD, E.C. and W.A. WALKER, *Hypothesis: inappropriate colonization of the premature intestine can cause neonatal necrotizing enterocolitis.* The FASEB Journal, 2001. **15**(8): p. 1398-1403.
45. Armstrong, J. and J.J.J.T.L. Reilly, *Breastfeeding and lowering the risk of childhood obesity.* 2002. **359**(9322): p. 2003-2004.
46. Lucas, A., et al., *Breast milk and subsequent intelligence quotient in children born preterm.* 1992. **339**(8788): p. 261-264.
47. Cong, X., et al., *Gut microbiome developmental patterns in early life of preterm infants: impacts of feeding and gender.* PloS one, 2016. **11**(4): p. e0152751.
48. Balmer, S. and B. Wharton, *Diet and faecal flora in the newborn: breast milk and infant formula.* Archives of Disease in Childhood, 1989. **64**(12): p. 1672-1677.
49. Martín, R.o., et al., *The commensal microflora of human milk: mew perspectives for food bacteriotherapy and probiotics.* Trends in Food Science & Technology, 2004. **15**: p. 121-127.
50. Harmsen HJ, W.-V.A., Raangs GC, et al. , *Analysis of intestinal flora development in breast-fed and formula-fed in- fants by using molecular identification and detection methods.* J Pediatr Gastroenterol Nutr, 2000. **30**(1): p. 61-67.
51. Gavin, A., & Ostovar, K., *Microbiological characterization of human milk.* Journal of Food Protection, 1977. **40**: p. 614-616.
52. Heikkila¨, M.P., & Saris, P. E. J., *Inhibition of Staphylococcus aureus by the commensal bacteria of human milk.* Journal of Applied Microbiology, 2003. **95**: p. 471-478.
53. West, P.A., Hewitt, J. H., & Murphy, O. M., *The influence of methods of collection and storage on the bacteriology of human milk.* Journal of Applied Bacteriology, 1979. **46**: p. 269-277.
54. Cohen-Wolkowiez, M., et al., *Early and late onset sepsis in late preterm infants.* The Pediatric infectious disease journal, 2009. **28**(12): p. 1052.
55. Tesini, B. *Neonatal Sepsis*. Pediatrics 2018 July 2018; Available from: https://www.merckmanuals.com/professional/pediatrics/infections-in-neonates/neonatal-sepsis.
56. Fouhy, F., et al., *High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin.* Antimicrobial agents and chemotherapy, 2012. **56**(11): p. 5811-5820.

57. Wandro, S., et al., *The Microbiome and Metabolome of Preterm Infant Stool Are Personalized and Not Driven by Health Outcomes, Including Necrotizing Enterocolitis and Late-Onset Sepsis.* mSphere, 2018. **3**(3): p. e00104-18.
58. Bonnemaison, E., et al., *Comparison of fecal flora following administration of two antibiotic protocols for suspected maternofetal infection.* Neonatology, 2003. **84**(4): p. 304-310.
59. Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.* Proceedings of the National Academy of Sciences, 2010. **107**(26): p. 11971-11975.
60. Biasucci, G., et al., *Cesarean delivery may affect the early biodiversity of intestinal bacteria.* The Journal of nutrition, 2008. **138**(9): p. 1796S-1800S.
61. Penders, J., et al., *Factors influencing the composition of the intestinal microbiota in early infancy.* Pediatrics, 2006. **118**(2): p. 511-521.
62. Adlerberth, I., Lindberg, E., Aberg, N., Hesselmar, B., Saalman, R., Strannegard, I.L., et al., *Reduced enterobacterial and increased staphylococcal colonization of the infantile bowel: an effect of hygienic lifestyle.* Pediatric Research, 2006. **59**: p. 96-101.
63. Hill, C.J., et al., *Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort.* Microbiome, 2017. **5**(1): p. 4.
64. Breitbart, M., et al., *Viral diversity and dynamics in an infant gut.* Research in microbiology, 2008. **159**(5): p. 367-373.
65. Minot, S., et al., *Rapid evolution of the human gut virome.* Proceedings of the National Academy of Sciences, 2013. **110**(30): p. 12450-12455.
66. Kapusinszky, B., P. Minor, and E. Delwart, *Nearly constant shedding of diverse enteric viruses by two healthy infants.* Journal of clinical microbiology, 2012. **50**(11): p. 3427-3434.
67. Reyes, A., et al., *Gut DNA viromes of Malawian twins discordant for severe acute malnutrition.* Proceedings of the National Academy of Sciences, 2015. **112**(38): p. 11941-11946.
68. Pannaraj, P., et al., *Shared and distinct features of human milk and infant stool viromes.* Frontiers in microbiology, 2018. **9**: p. 1162.
69. Minot, S., et al., *The human gut virome: inter-individual variation and dynamic response to diet.* Genome research, 2011. **21**(10): p. 1616-1625.
70. Stewart, J.A., V.S. Chadwick, and A. Murray, *Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children.* Journal of medical microbiology, 2005. **54**(12): p. 1239-1242.
71. Reyes, A., et al., *Viruses in the faecal microbiota of monozygotic twins and their mothers.* Nature, 2010. **466**(7304): p. 334.
72. McCann, A., et al., *Viromes of one year old infants reveal the impact of birth mode on microbiome diversity.* PeerJ, 2018. **6**: p. e4694.
73. Gregory, K.E., et al., *Influence of maternal breast milk ingestion on acquisition of the intestinal microbiome in preterm infants.* Microbiome, 2016. **4**(1): p. 68.
74. Cong, X., et al., *Influence of Infant Feeding Type on Gut Microbiome Development in Hospitalized Preterm Infants.* Nursing research, 2017. **66**(2): p. 123.
75. Parra-Llorca, A., et al., *Preterm Gut Microbiome Depending on Feeding Type: Significance of Donor Human Milk.* Frontiers in microbiology, 2018. **9**: p. 1376.

76. Gewolb, I.H., et al., *Stool microflora in extremely low birthweight infants.* 1999. **80**(3): p. F167-F173.
77. Stewart, C., et al., *The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection.* 2012. **101**(11): p. 1121-1127.
78. Schwiertz, A., et al., *Development of the intestinal bacterial composition in hospitalized preterm infants in comparison with breast-fed, full-term infants.* 2003. **54**(3): p. 393.
79. Gohl, D.M., et al., *Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies.* Nature biotechnology, 2016. **34**(9): p. 942.
80. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.
81. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nature methods, 2015. **12**(1): p. 59.
82. Huson, D.H., et al., *MEGAN analysis of metagenomic data.* Genome research, 2007. **17**(3): p. 377-386.
83. Bolyen, E., et al., *QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science*. 2018, PeerJ Preprints.
84. Edgar, R.C., *UPARSE: highly accurate OTU sequences from microbial amplicon reads.* Nature methods, 2013. **10**(10): p. 996.
85. Martín, R., et al., *Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women.* Research in microbiology, 2007. **158**(1): p. 31-37.
86. Jost, T., et al., *Vertical mother–neonate transfer of maternal gut bacteria via breastfeeding.* Environmental microbiology, 2014. **16**(9): p. 2891-2904.
87. Jiménez, E., et al., *Metagenomic analysis of milk of healthy and mastitis-suffering women.* Journal of Human Lactation, 2015. **31**(3): p. 406-415.
88. Hoffmann, C., et al., *Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents.* PloS one, 2013. **8**(6): p. e66019.
89. Organization, W.H., *Mastitis: Causes and Management, WHO/FCH/CAH/00.13.* 2000.
90. Martin, V., et al., *Sharing of Bacterial Strains Between Breast Milk and Infant Feces.* Journal of Human Lactation, 2012. **28**(1): p. 36-44.
91. Behari, P., et al., *Transmission of methicillin-resistant Staphylococcus aureus to preterm infants through breast milk.* Infection Control and Hospital Epidemiology, 2004. **25**(9): p. 778-780.
92. Eickhoff, T.C., et al., *Neonatal sepsis and other infections due to group B beta-hemolytic streptococci.* New England Journal of Medicine, 1964. **271**(24): p. 1221-1228.
93. Stoll, B.J., et al., *Changes in pathogens causing early-onset sepsis in very-low-birth-weight infants.* New England Journal of Medicine, 2002. **347**(4): p. 240-247.
94. Barton, L.L., R.D. Feigin, and R. Lins, *Group B beta hemolytic streptococcal meningitis in infants.* The Journal of pediatrics, 1973. **82**(4): p. 719-723.
95. Le Doare, K. and B. Kampmann, *Breast milk and Group B streptococcal infection: vector of transmission or vehicle for protection?* Vaccine, 2014. **32**(26): p. 3128-3132.

96. Maraki, S. and I.S. Papadakis, *Rothia mucilaginosa pneumonia: a literature review.* Infectious Diseases, 2015. **47**(3): p. 125-129.
97. Wang, M., et al., *Fecal Microbiota Composition of Breast-Fed Infants Is Correlated With Human Milk Oligosaccharides Consumed.* Journal of Pediatric Gastroenterology and Nutrition, 2015. **60**(6): p. 825-833.
98. Jost, T., et al., *Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches.* British Journal of Nutrition, 2013. **110**(7): p. 1253-1262.
99. Biesbroek, G., et al., *The Impact of Breastfeeding on Nasopharyngeal Microbial Communities in Infants.* American Journal of Respiratory and Critical Care Medicine, 2014. **190**(3): p. 298-308.
100. Grabherr, M.G., et al., *Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data.* Nature biotechnology, 2011. **29**(7): p. 644.